

Volatility Decomposition and Estimation in Time-Changed Price Models

Rainer Dahlhaus*

Institute of Applied Mathematics, Heidelberg University
Im Neuenheimer Feld 205, 69120

Heidelberg, Germany

E-mail: dahlhaus@statlab.uni-heidelberg.de

and

Sophon Tunyavetchakit

Institute of Applied Mathematics, Heidelberg University
Im Neuenheimer Feld 205, 69120

Heidelberg, Germany

E-mail: s.tunyavetchakit@uni-heidelberg.de.

May 10, 2016

Abstract

The usage of a spot volatility estimate based on a volatility decomposition in a time-changed price-model according to the trading times is investigated. In this model clock-time volatility splits up into the product of tick-time volatility and trading intensity, which both can be estimated from data and contain valuable information. By inspecting these two curves individually we gain more insight into the cause and structure of volatility. Several examples are provided where the tick-time volatility curve is much smoother than the clock-time volatility curve meaning that the major part of fluctuations in clock-time volatility is due to fluctuations of the trading intensity. Since microstructure noise only influences the estimation of the (smooth) tick-time volatility curve, the findings lead to an improved pre-averaging estimator of spot volatility. This is reflected by a better rate of convergence of the estimator. The asymptotic properties of the estimators are derived by an infill asymptotic approach.

Keywords: Time-changed Brownian motion, tick-time volatility, high-frequency transaction data, market microstructure noise, pre-averaging method.

*This work was supported by the Deutsche Forschungsgemeinschaft, RTG 1953. The authors thank Per Mykland and Lan Zhang for very helpful comments.

1. INTRODUCTION

The estimation of volatility for high-frequency data under microstructure noise has been extensively studied during recent years - see Aït-Sahalia and Jacod (2014) for an overview. The majority of this work has been carried out in the framework of diffusion models. In this paper we focus on the estimation of spot volatility, and, contrary to the majority of previous research, on a time-changed price-model based on trading times. In this model volatility splits up into the product of two identifiable curves, namely tick-time volatility and trading intensity. The main methodological and theoretical contributions of this paper are the introduction and the theoretical investigation of a volatility estimate based on this volatility decomposition, and the proof that this estimator can outperform the classical (diffusion model based) estimators in terms of the rate of convergence.

For the estimation of tick-time volatility under microstructure noise we have to adapt an estimator from diffusion models to our situation which can cope successfully with microstructure noise in a high-frequency situation. Many noise-robust estimators have been introduced in the literature. In Zhang et al. (2005) the combination of two different timescales is used to construct a consistent estimator for the integrated volatility. This idea is extended later to the multi-timescale estimator that archives the optimal rate of convergence $n^{-1/4}$ - cf. Zhang (2006). Barndorff-Nielsen et al. (2008) suggest a flat-top kernel-type estimator, called realized kernel, which combines different lags of autocovariances to eliminate the effect of microstructure noise. Recently, an estimator presented by Reiß (2011) and Bibinger and Reiß (2014) has received attention since their estimator is asymptotically efficient. In this work, we apply the pre-averaging technique, which was introduced by Podolskij and Vetter (2009) and later extended by Jacod et al. (2009), in order to construct a noise-robust estimate for the tick-time volatility in our time-changed model. However, it is also possible to adapt most of the other methods to the model of this paper. An estimator based on particle filtering in a nonlinear microstructure noise model has been discussed in Dahlhaus and Neddermeyer (2013).

Time-changed price-models were first investigated in Clark (1973) in connection with finance. In his work, the volume of trades is suggested to be a subordinator of a Brownian motion in order to recover the normality of the distribution of cotton future prices. Afterwards a relationship between asset returns, price fluctuation, and market activities measured by trading volume and numbers of transactions is extensively discussed. Ané and Geman (2000) conclude that the number of trades explains the volatility change better than their volume, so they recover the normality of asset returns through this stochastic time change in high-frequency data; see also Jones et al. (1994), Plerou et al. (2001) and Gabaix et al. (2003) for more detailed discussions of this correlation. Due to various mathematical tools, the time-changed Brownian motion is attractive and tractable to study arbitrage-free asset returns, which are shown to be semimartingales (see e.g. Delbaen and Schachermayer 1994). Indeed, having a class of time-changed Brownian motion is satisfactory since it is as large as a class of semimartingale; see Monroe (1978). In recent years, other time change models have been extensively studied, especially a time-changed Lévy process which allows for a more complex structure in the price models coping with some stylized-effect emerging in the real market; for details refer to Carr et al. (2003) and Carr and Wu (2004), and to Belomestny (2011) for a statistical treatment of this kind of models.

The article is organized as follows. Section 2 contains an introduction to the model, the volatility decomposition, and the estimates with a discussion of the implications for applications. Here, the volatility decomposition is proven under a general setup. In Section 3 we investigate the asymptotic properties of the estimates by means of an infill asymptotic approach constructed by time-rescaling. The asymptotic results are compared in Table 1 which shows the advantage of using the volatility decomposition. In particular the results show that the rate of convergence of classical estimators can be outperformed within the model of this paper. Section 4 contains some concluding remarks.

The details of the data analysis and the proofs can be found in the Appendix. If not otherwise stated, all equalities and inequalities of random expressions are in an almost sure sense.

2. THE VOLATILITY DECOMPOSITION

As motivated in the introduction we use instead of the classical semimartingale model a diffusion model subordinated by transaction time, for example the time-changed Brownian motion

$$dX_t = \sigma_t dW_{N_t} \quad \text{for } t \in [0, T] \quad (1)$$

with N_t being a point process with intensity λ_t reflecting the accumulated number of transactions up to time t . In it's simplest form σ_t and λ_t are deterministic and $W(\cdot)$ and N are independent. In a more general model σ_t and λ_t are stochastic processes depending on the past of X_t and N_t , the independence of $W(\cdot)$ and N may be replaced by some martingale-structure, and $W(\cdot)$ may be non-Gaussian. We include microstructure noise into our considerations - for example the asymptotic properties of our estimates are derived under the assumption of additive i.i.d. noise

$$Y_{t_i} = X_{t_i} + \varepsilon_i \quad \text{for } i = 1, \dots, N_T, \quad (2)$$

where t_i are the trading times.

The focus of this paper is the estimation of spot volatility for financial transactions which is not the function σ_t^2 from the above model. Even more the meaning of σ_t^2 is different from the meaning of $\sigma_{clock}^2(t)$ in the classical “clock-time” diffusion model (say $dX_t = \sigma_{clock}(t) dW_t$). We therefore start with a model-independent definition of spot volatility and clarify the relation to σ_t in the different models. Let $(\mathcal{F}_t)_{t \geq 0}$ be an increasing sequence of σ -algebras; roughly speaking, it represents the information available up to and including time t . We define

$$\text{vola}_t^2 := \lim_{\Delta t \rightarrow 0} \frac{\mathbb{E}[(\mathbf{X}(t + \Delta t) - \mathbf{X}(t))^2 | \mathcal{F}_t]}{\Delta t}.$$

In the classical diffusion model, vola_t^2 does not depend on the point process and we have under the assumption that $\sigma_{clock}(t)$ is a right-continuous process with left-limits adapted to \mathcal{F}_t : $\text{vola}_t^2 = \sigma_{clock}^2(t)$. Therefore, we use $\sigma_{clock}(t)$ in this paper as a synonym for vola_t , i.e. we define

$$\sigma_{clock}^2(t) := \text{vola}_t^2.$$

In the transaction-time model of this paper we prove below that $\sigma_{clock}^2(t) = \sigma_t^2 \cdot \lambda_t$. We first set down the assumptions for this result. To understand our assumptions, note that in (1) we do not use the whole process $W(\cdot)$ but only the increments $U_i := W(N_{t_i}) - W(N_{t_{i-1}})$ which we now assume to be a martingale difference sequence.

2.1 Assumption *The X_{t_i} at observation times t_i follow the model $X_{t_i} = X_{t_{i-1}} + \sigma_{t_i} U_i$ where the t_i are the arrival times of a point process N_t . We assume that there exists a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, where \mathcal{F}_0 includes all null sets and the filtration $(\mathcal{F}_t)_{t \geq 0}$ is right-continuous, such that*

- i) N_t is a point process admitting an \mathcal{F}_t -intensity λ_t (as in Definition D7 of Brémaud 1981); in particular λ_t is an \mathcal{F}_t -progressive process and N_t is adapted to \mathcal{F}_t ;
- ii) σ_t^2 is a non-negative \mathcal{F}_t -predictable process; in particular σ_t^2 is \mathcal{F}_{t-} -measurable;

iii) U_i is \mathcal{F}_{t_i} -measurable, for each i , with

$$\mathbb{E} \left[U_i \mid \mathcal{F}_{t_i-} \right] = 0 \quad \text{and} \quad \mathbb{E} \left[U_i^2 \mid \mathcal{F}_{t_i-} \right] = 1.$$

Note that no other condition on the distribution of U_i is required. Examples for processes which fulfill these assumptions are given at the end of this chapter. The smallest filtration which satisfies the above conditions is

$$\mathcal{F}_t = \sigma(\{N_s : s \leq t\}, \{\lambda_s : s \leq t\}, \{\sigma_s : s \leq t\}, \{U_{N_s} : s \leq t\}).$$

2.2 Proposition Suppose Assumption 2.1 holds. If σ_t and λ_t are continuous processes we have

$$\sigma_{clock}^2(t) = \sigma_t^2 \cdot \lambda_t. \quad (3)$$

The proof can be found in Appendix A.2.

Thus, in the transaction-time model, the volatility can be decomposed into the product of two curves which both can be identified from the data. There exists an intuitive interpretation of this decomposition in that the formula reflects the change of time unit. Note that σ_t^2 can be seen as the volatility per transaction and $\sigma_{clock}^2(t)$ as the volatility per calendar time unit. The formula then says that “volatility per time unit is equal to volatility per transaction multiplied by the average number of transactions per time unit”.

Both curves σ_t^2 and λ_t can be estimated by various estimates $\hat{\sigma}^2(t)$ and $\hat{\lambda}(t)$. We use these estimates in two ways:

(i) to construct an alternative estimator of $\sigma_{clock}^2(t)$ via

$$\tilde{\sigma}_{clock}^2(t) := \hat{\sigma}^2(t) \cdot \hat{\lambda}(t);$$

(ii) to look at the two curves individually in order to gain more insight about the cause and the structure of volatility.

In this paper we use a kernel estimate for λ_t (with $\int_{\mathbb{R}} \mathfrak{K}(x) dx = 1$ and $\mathfrak{K}(x) = 0$ for $|x| \geq 1$ - the same for the kernels k and K from below)

$$\hat{\lambda}(t_0) := \frac{1}{M} \sum_{i=1}^{N_T} \mathfrak{K}\left(\frac{t_i - t_0}{M}\right) \quad (4)$$

where $t_0 \in (0, T)$, and in order to handle microstructure noise, the pre-averaging technique of Podolskij and Vetter (2009), extended by Jacod et al. (2009), for the estimation of σ_t^2 (adapted to the present model)

$$\begin{aligned} \hat{\sigma}_{pavg}^2(t_0) := & \frac{1}{mH} \frac{1}{g_2} \sum_{i=i_0-m}^{i_0+m} k\left(\frac{i-i_0}{m}\right) (\overline{\Delta Y}_{t_i})^2 \\ & - \frac{1}{2mH} \frac{\sum_{l=1}^{H-1} h^2(\frac{l}{H})}{g_2} \sum_{i=i_0-m}^{i_0+m} k\left(\frac{i-i_0}{m}\right) (Y_{t_i} - Y_{t_{i-1}})^2 \end{aligned}$$

where $i_0 := \inf \{i : t_i \geq t_0\}$,

$$\overline{\Delta Y}_{t_i} := \sum_{l=1}^{H-1} g\left(\frac{l}{H}\right) (Y_{t_{i+l}} - Y_{t_{i+l-1}})$$

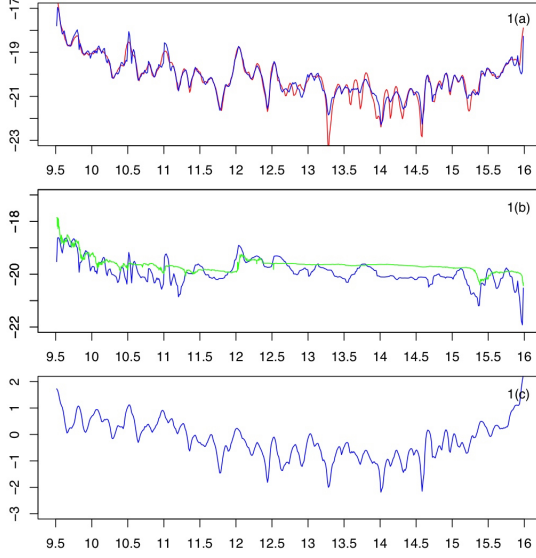


Figure 1: Volatility of MSFT on April 1, 2014 based on 25,198 transactions, $M=200$: clock-time volatility(a), transaction-time volatility(b), trading intensity(c).

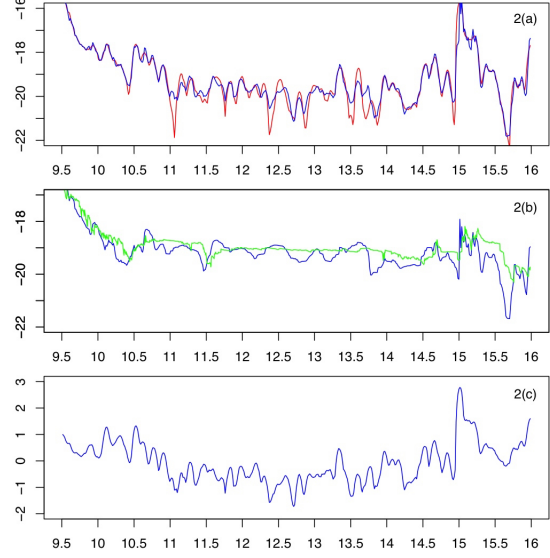


Figure 2: Volatility of GM on April 1, 2014 based on 31,044 transactions, $M=200$: clock-time volatility(a), transaction-time volatility(b), trading intensity(c).

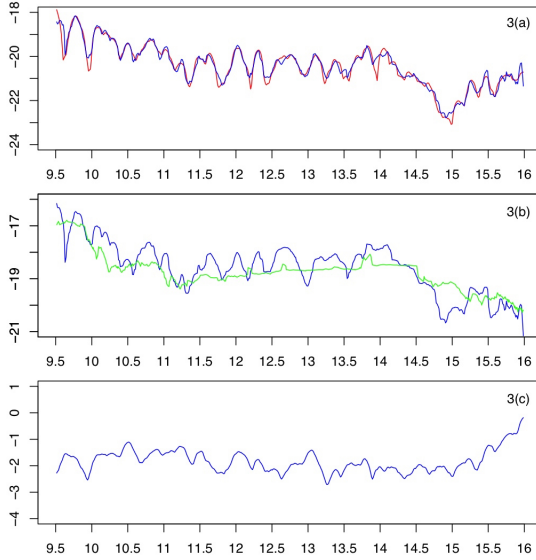


Figure 3: Volatility of HON on April 1, 2014 based on 4,162 transactions, $M=300$: clock-time volatility(a), transaction-time volatility(b), trading intensity(c).

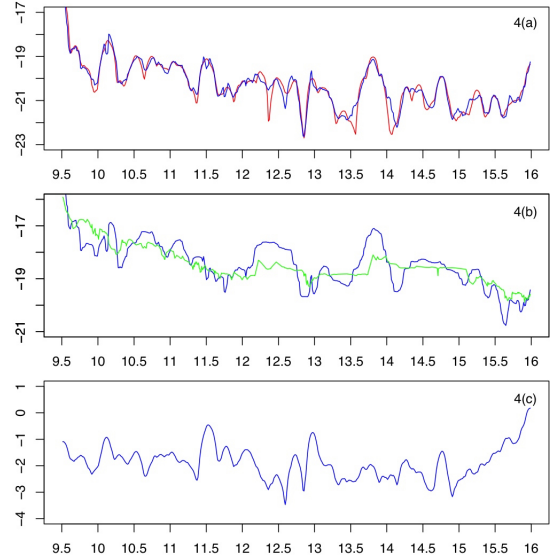


Figure 4: Volatility of NKE on April 1, 2014 based on 4,341 transactions, $M=300$: clock-time volatility(a), transaction-time volatility(b), trading intensity(c).

Figure 1–Figure 4: The first row in each plot shows the log of clock-time volatility $\log \hat{\sigma}_{clock,pavg}^2(t)$ (red) and $\log \hat{\sigma}_{clock,pavg}^2(t)$ (blue), the second row in each plot shows the log of tick-time volatility $\log \hat{\sigma}_{pavg}^2(t)$ (blue) and $\log \hat{\sigma}^2(t)$ (green), and the third row the log trading intensity $\log \hat{\lambda}(t)$ (blue). The blue estimators from row (b) and (c) sum up to the blue estimator in (a).

and $h(l/H) := g((l+1)/H) - g(l/H)$. H is the smoothing parameter in the pre-averaging step (for more details see Section 3). This leads to the new alternative clock-time estimator based on the volatility decomposition (3)

$$\tilde{\sigma}_{clock,pavg}^2(t_0) := \hat{\sigma}_{pavg}^2(t_0) \cdot \hat{\lambda}(t_0) \quad (5)$$

while the “classical” pre-averaging clock-time volatility estimator is

$$\begin{aligned} \hat{\sigma}_{clock,pavg}^2(t_0) = & \frac{1}{MH} \frac{1}{g_2} \sum_{i=1}^{N_T} K\left(\frac{t_i - t_0}{M}\right) (\overline{\Delta Y}_{t_i})^2 \\ & - \frac{1}{2MH} \frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2} \sum_{i=1}^{N_T} K\left(\frac{t_i - t_0}{M}\right) (Y_{t_i} - Y_{t_{i-1}})^2. \end{aligned}$$

In Figures 1–4 we have applied both estimates to high-frequency data, approximately 4,000–30,000 transactions per day, from the NASDAQ stock exchange (MSFT = Microsoft, GM = General Motors, HON = Honeywell, NKE = Nike) - more details about the data can be found in Appendix A.1. The first row always shows the logarithm $\log \tilde{\sigma}_{clock,pavg}^2(t)$ (red) and $\log \hat{\sigma}_{clock,pavg}^2(t)$ (blue), the second row $\log \hat{\sigma}_{pavg}^2(t)$ (blue) and another tick-time volatility estimator $\log \hat{\sigma}^2(t)$ (green) from Dahlhaus and Neddermeyer (2013) - see the discussion below, and the third row the log trading intensity $\log \hat{\lambda}(t)$ (blue). More details about the estimators can also be found in Appendix A.1. Due to the additive relation

$$\log \tilde{\sigma}_{clock,pavg}^2(t) = \log \hat{\sigma}_{pavg}^2(t) + \log \hat{\lambda}(t),$$

the blue curves in the second and third row sums up to the blue estimator in the first row. From the figures it can be seen that

(i) row (a) shows that the new estimator of this paper based on relation (3) (blue curve) nicely coincides with the classical clock-time estimator (red). This blue estimator is the sum of the blue estimators in row (b) and (c) (in log scale);

(ii) the tick-time volatility estimator $\log \hat{\sigma}_{pavg}^2(t)$ in row (b) is in general smoother than the clock-time estimator in row (a) and $\log \hat{\lambda}(t)$ in row (c) - i.e. the fluctuation of trading intensity in row (c) is the major source of fluctuation of clock-time volatility in row (a). This effect is quite clear for the high-liquid stocks MSFT and GM in Figures 1–2 respectively (where $M = 200$; $m = 215$ and $m = 265$ respectively - see Section A.1). For the less liquid stocks HON and NKE in Figures 3–4 (where $M = 300$; $m = 53$ and $m = 55$) the effect is less visible - in our opinion due to the considerably lower tick time bandwidth m .

(iii) the decomposition allows to a certain extent to determine the source of volatility changes: for example the peak in 4(a) at time 13.7 is due to a peak of tick-time volatility (i.e. most likely due to some company related news) while the peak in 2(a) at time 15.1 is mainly due to a peak of trading intensity (i.e. most likely due to some general - not company related - news). Similarly the decrease of volatility in 3(a) after 14.1 is company related;

(iv) in particular the curves in 1(a) and 2(a) exhibit the typical U-shape over the trading day. It is notable that this U-shape is mainly a feature of the trading intensity in row (c). In Fig. 3 and 4 the U-shape in (c) is compensated by a decrease of tick-time volatility in (b) at the end of the trading day.

In most of our examples tick-time volatility in row (b) is considerably smoother than trading intensity in row (c). Beyond interpretation this has also an important consequence for estimation:

microstructure noise only affects the smoother curve (b) and not (c), i.e. coping with microstructure noise becomes easier since we may choose a larger bandwidth with effectively more data than with the classical estimator (red curve in (a)). Mathematically this is reflected in a higher rate of convergence of the estimate (see Section 3 and in particular Table 1) - in particular we may even outperform the lower bound in diffusion models (provided that the time-rescaled model of this paper is correct).

There is an open issue about the quality of the pre-averaging estimator in (b) (a detailed investigation of this problem is beyond the scope of this paper - we just mention it briefly): The green estimator in (b) permanently is even smoother than the blue pre-averaging estimator. This estimator is completely different: it uses a nonlinear microstructure noise model with particle filtering and adaptive bandwidth selection - see Dahlhaus and Neddermeyer (2013) for details. We have studied the behavior of both estimators by a simulation in the case where (i) the true tick-time volatility is constant (Figure 5) - emulating the case where the true curve is similar to the green estimate in 1(b), and in the case where (ii) the true tick-time volatility is oscillating (Figure 6) - emulating the case where the true curve is similar to the blue estimate in 3(b). As the microstructure noise model we have chosen in the simulations additive noise plus rounding. The plots in Figures 5 and 6 indicate that the green estimator resembles in particular a constant curve in a better way, and is not close to constant if the true curve is not constant, meaning that the true unknown tick-time volatility-curve in Figures 1(b)–4(b) is likely to be closer to the green curve than to the blue curve. In particular this confirms that tick-time volatility is usually smoother than clock-time volatility and trading intensity.

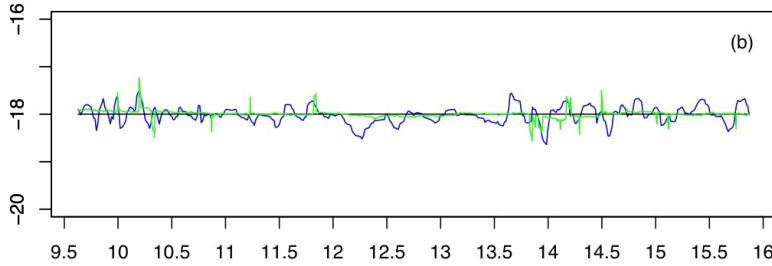


Figure 5: Tick-time volatility estimates $\log \hat{\sigma}_{pavg}^2(t)$ (blue) and $\log \hat{\sigma}^2(t)$ (green) for simulated transactions with true constant volatility (black)

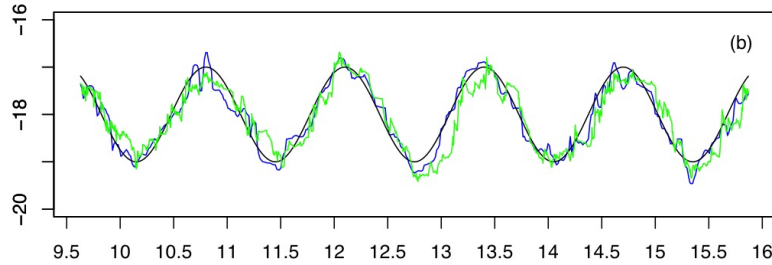


Figure 6: Tick-time volatility estimates $\log \hat{\sigma}_{pavg}^2(t)$ (blue) and $\log \hat{\sigma}^2(t)$ (green) for simulated transactions with true oscillating volatility (black)

We investigate the properties of the alternative clock-time estimator $\tilde{\sigma}_{clock}^2(t) := \hat{\sigma}^2(t) \cdot \hat{\lambda}(t)$ in the next section. At the end of this section we give some examples where Assumption 2.1 is fulfilled.

2.3 Example (i) The simplest case is the model in (1) where σ_t and λ_t are deterministic and $W(\cdot)$ and N_\cdot are independent. Even in this case the derivation of asymptotic results for the estimates is non-standard since the classical asymptotic setting cannot be applied. We therefore introduce a type of infill asymptotics for this setup in the next section.

(ii) The more general model allowing for stochastic parameters is $dX_t = \sigma_t dL_t$, where L_t is a pure jump process of the form $\sum_{i=1}^{N_t} U_i$. The point process N_t and the sequence of innovations U_i need to satisfy the conditions given in Assumption 2.1, i.e. N_t has an \mathcal{F}_t -intensity λ_t and U_i has a martingale difference structure. Thus, L_t can be seen as a generalization of a Lévy process without a diffusion part. The leverage effect between all processes can be constructed, for example, by setting $\lambda_t := \alpha_t^2$ with

$$d\alpha_t = a_t dB_t + a'_t dB'_t \quad \text{and} \quad d\sigma_t = b_t dB_t + b'_t dB''_t,$$

where B_t , B'_t , and B''_t are three different \mathcal{F}_t -Brownian motions; a_t , a'_t , b_t and b'_t are \mathcal{F}_t -adapted processes. The dependence between the price process and the intensity (also the tick-time volatility) lies in the process a_t (also in a'_t , b_t and b'_t). For instance, we could model a_t by $da_{t+} = a_t^* dX_t$ where a_t^* is an adapted predictable process.

(iii) It is interesting that GARCH-models “almost” fit into this framework: For example let N_t be a doubly stochastic Poisson process with right-continuous λ_t and

$$\sigma_s^2 := a_0 + a \{X_{t_{j(s)}} - X_{t_{j(s)-1}}\}^2 + b\sigma_{t_{j(s)}}^2 \in \mathcal{F}_{s-},$$

where $j(s) := \max \{j : t_j < s\}$ and a_0, a, b are positive constants. Clearly, this volatility function σ_t^2 is a left-continuous \mathcal{F}_t -adapted step function. Since it is not right continuous we obtain instead of (3) with the same proof

$$\sigma_{clock}^2(t+) = \sigma_{t+}^2 \cdot \lambda_t$$

(note that $(\mathcal{F}_t)_{t \geq 0}$ is right-continuous).

(iv) In the spirit of the last example we may construct more complex examples where λ_t is also a left-continuous \mathcal{F}_t -adapted depending both on past arrival times of $N(t)$ (e.g. via a similar structure as in Hawkes-models) and on past log-prices. σ_t may in addition to the GARCH-structure from above also depend on the intensity of the point process. In that way we may explicitly model the dependence between the price process and the trading intensity.

3. INFILL ASYMPTOTICS

For asymptotic investigations of the estimators (e.g. of $\tilde{\sigma}_{clock,avg}^2(t_0)$ in (5)) we need a setup where the number of trading times increases in the neighborhood of t_0 . A first idea would be to use a point process with intensity $n\lambda(t)$ where $n \rightarrow \infty$ (cf. Bibinger et al. 2016, eq. (1.2)). Instead we use here an infill asymptotic approach similar to nonparametric regression or locally stationary time series (Dahlhaus 1997). The reason for doing so is explained in Remark 3.2 below.

We assume that the asset log-price process is of the form

$$dX_{t,T} = \sigma \left(\frac{t}{T} \right) \frac{1}{\sqrt{T}} dW_{N_{t,T}}, \quad \text{for } t \in [0, T]. \quad (6)$$

W_\cdot is a standard Brownian motion and $N_{t,T}$ is a nonhomogeneous Poisson process (NHPP) with a continuous non-negative real-valued intensity function $\lambda(t/T)$; $W_{N_{t,T}}$ is therefore a time-changed

Brownian motion. In this model $\sigma^2(\cdot)$ is a real-valued deterministic continuous function called *tick-time volatility*, since it responds to price variation from one trade to the next. By $t_{i,T} := \inf \{t : N_{t,T} \geq i\}$ we denote the arrival times of the point process. In order to not complicate the notation we often avoid the double subscript, i.e. we set $t_i := t_{i,T}$ and $N_T := N_{T,T}$. We include microstructure noise into our model (cf. Zhang et al. 2005; Bandi and Russell 2008); for simplicity we assume that the observations follow the linear model

$$Y_{t_i,T} = X_{t_i,T} + \varepsilon_i \quad \text{for } i = 1, \dots, N_T, \quad (7)$$

where ε_i is assumed to be i.i.d. and independent of $X_{t,T}$ with

$$\mathbb{E}[\varepsilon_i] = 0, \quad \text{Var}[\varepsilon_i] = \omega^2 < \infty \quad \text{and} \quad \text{Var}[\varepsilon_i^2] = \theta\omega^4, \quad \text{for } \theta \in \mathbb{R}^+.$$

In fact, it has been empirically shown that the assumption about the independence of the noise is justifiable for high-frequency intraday data with transaction sampling, but not for other sampling schemes such as quotation sampling, business-time sampling, 1-minute sampling, etc. (Hansen and Lunde 2006; Griffin and Oomen 2008). Throughout this section we will work under the following conditions.

3.1 Assumption i) *The processes $N_{\cdot,T}$ and W_{\cdot} are independent;*

ii) *$\sigma^2(\cdot)$ is deterministic, lies in the Hölder class $\mathcal{C}^{m,\gamma}[0,1]$, for $m = 0, 1, 2$ and $0 < \gamma < 1$, and is bounded away from zero uniformly in u ;*

iii) *$\lambda(\cdot)$ is deterministic, lies also in the Hölder class $\mathcal{C}^{m',\gamma'}[0,1]$, for $m' = 0, 1, 2$ and $0 < \gamma' < 1$, and is bounded away from zero uniformly in u .*

To recall the definition of the Hölder class, $f \in \mathcal{C}^{m,\gamma}[0,1]$ for $0 < \gamma < 1$ and $m \geq 0$, if

$$|f^{(m)}(x + \delta) - f^{(m)}(x)| \leq C \cdot |\delta|^\gamma, \quad \text{for } |\delta| \rightarrow 0 \text{ and a constant } C.$$

For model (6), the variance of the price increment over $[t_o, t_o + bT]$ is given by

$$\mathbb{E}[X_{t_o+bT,T} - X_{t_o,T}]^2 = \int_{u_o}^{u_o+b} \sigma^2(u) \lambda(u) du,$$

where $u_o := t_o/T$. This implies

$$\sigma_{clock}^2(u_o) := \lim_{b \rightarrow 0} \frac{1}{b} \mathbb{E}[X_{t_o+bT,T} - X_{t_o,T}]^2 = \sigma^2(u_o) \cdot \lambda(u_o), \quad (8)$$

i.e. we obtain the same volatility decomposition as in Proposition 2.2 also in the rescaled model.

Many common parametric volatility models in the literature, such as the Heston, GARCH, CIR, etc., are stochastic processes driven by a Brownian motion. Hence their realizations, being non-differentiable, lie in this class with $m = 0$ and $\gamma < 1/2$. In our setting the volatility curve $\sigma_{clock}^2(\cdot)$ acts like the one that is driven by a Brownian motion whenever either of the components of (8) lies in the Hölder class $\mathcal{C}^{0,\gamma}$ with $\gamma < 1/2$, while the other component is allowed to be smoother (as a consequence, one may choose larger bandwidth for estimating this curve - see below).

Recently other parametric volatility models relying on a fractional Brownian motion have also been applied in price modeling. In fact, almost-all trajectories of this process with Hurst index $H \in (0,1)$ lie in the Hölder class with $m = 0$ and $\gamma < H$. From this viewpoint the Hölder smoothness class seems to be an appropriate choice of smoothness class for our parameter curves.

3.2 Remark i) As mentioned above, an alternative model to (6) seems to be a non-rescaled price model with intensity $n\lambda(t)$ where $n \rightarrow \infty$. Both approaches however are similar: In this alternative model a segment about t_o of length b (bandwidth of kernel estimators) would contain in the average $b \times n\lambda(t_o)$ data points while in the rescaled model of this paper a segment about $u_o := t_o/T$ of length b would contain in the average $bT \times \lambda(u_o)$ data points. Therefore, we conjecture that the two asymptotic approaches are equivalent for $n = T$. The reason for choosing the rescaled infill asymptotic model is that this approach seems to be more flexible towards future generalizations where also dependence between the price process and the trading intensity is included (which seems to be difficult if the trading intensity converges to infinity).

ii) Another aspect which looks strange at first sight is the factor $1/\sqrt{T}$ in the price model (6) meaning that the ‘true tick-time volatility’ is not $\sigma^2(t/T)$ but $\sigma^2(t/T)/T$. Similarly the ‘true clock-time volatility is $\sigma_{\text{clock}}^2(u_o)/T$, i.e. the decomposition formula still holds regardless of this rescaling. The main reason for introducing this factor is that then the influence of microstructure noise is asymptotically the same as in the classical case:

To be more precise, consider for example the simple realized volatility estimate $\hat{v} := \frac{1}{b} \sum_{t_i \in I} [Y_{t_i} - Y_{t_{i-1}}]^2$ on a segment I of length b in a model with constant $\sigma(\cdot) \equiv \sigma$, and trading times t_i from an independent Poisson process with time-constant $\lambda(\cdot) \equiv \lambda$. We then have:

- in the classical diffusion model $dX_t = \sigma dW_t$ with $I := [t_o, t_o + b]$, trading times $t_i = i/T$, or alternatively trading times t_i generated from a Poisson process with intensity $\lambda = T$:

$$\mathbb{E}[\hat{v}] = \mathbb{E} \frac{1}{b} \sum_{t_i \in I} [\sigma^2(t_i - t_{i-1}) + 2\omega^2] \approx \sigma^2 + 2T\omega^2;$$

- in the tick-time model (6) with factor $1/\sqrt{T}$ and $I := [t_o/T, t_o/T + b]$, $\lambda = 1$,

$$\mathbb{E}[\hat{v}] = \mathbb{E} \frac{1}{b} \sum_{t_i/T \in I} [\sigma^2/T + 2\omega^2] \approx \frac{1}{b} \sigma^2 b \lambda + 2\lambda T \omega^2 = \sigma^2 + 2T\omega^2;$$

- in the tick-time model (6) without factor $1/\sqrt{T}$ and $I := [t_o/T, t_o/T + b]$, $\lambda = 1$,

$$\mathbb{E}[\hat{v}] = \mathbb{E} \frac{1}{b} \sum_{t_i/T \in I} [\sigma^2 + 2\omega^2] \approx \sigma^2 \lambda T + 2\lambda T \omega^2 = \sigma^2 T + 2T\omega^2,$$

meaning that we had to replace \hat{v} by \hat{v}/T leading to $\mathbb{E}\hat{v} \approx \sigma^2 \lambda + 2\lambda \omega^2$, i.e. the microstructure noise term were asymptotically of a lower order than in the classical case.

Thus we need the factor $1/\sqrt{T}$ in model (6) in order to make the microstructure noise problem comparable to the classical case.

iii) The independence between $N_{\cdot,T}$ and W . enables us to make arguments conditional on the entire process $N_{\cdot,T}$ which greatly simplifies the proofs. In particular, for consecutive arrival times t_{j-1} and t_j we have

$$X_{t_j,T} - X_{t_{j-1},T} \stackrel{\text{law}}{=} \sigma \left(\frac{t_j}{T} \right) \frac{1}{\sqrt{T}} U_j,$$

since $W_{N_{t,T}}$ has the same law as $\sum_{i=1}^{N_{t,T}} U_i$, where U_i are i.i.d. normally distributed variables with zero mean and unit variance and independent of $N_{\cdot,T}$.

A leverage effect between W and $N_{\cdot,T}$ could be allowed for in order to explain the correlation between market activities and price processes; see the formation of general models in Assumptions 2.1. A thoroughly asymptotic investigation of this complex transaction-time model under those assumptions is beyond the scope of this paper.

We now investigate the properties of the estimates of Section 2. For the sake of clarity we redefine them in the new infill asymptotic approach. The transaction intensity $\lambda(\cdot)$ is estimated by

$$\hat{\lambda}(u_o) := \frac{1}{\mathfrak{b}T} \sum_{i=1}^{N_T} \mathfrak{K} \left(\frac{t_i - u_o T}{\mathfrak{b}T} \right) = \frac{1}{\mathfrak{b}T} \int_0^T \mathfrak{K} \left(\frac{t - u_o T}{\mathfrak{b}T} \right) dN_{t,T}, \quad (9)$$

where the bandwidth $\mathfrak{b} = \mathfrak{b}(T) \rightarrow 0$ and $\mathfrak{b}T \rightarrow \infty$ as $T \rightarrow \infty$. The kernel function \mathfrak{K} (and also K and k given below) satisfies the following conditions:

Condition (K) The kernel function $\mathfrak{K} : \mathbb{R} \rightarrow \mathbb{R}^+$ is a continuous, symmetric function such that $\mathfrak{K}(x) = 0$ for $|x| \geq 1$ and $\int_{\mathbb{R}} \mathfrak{K}(x) dx = 1$.

3.3 Theorem Let Assumption 3.1 holds, then $\hat{\lambda}(u_o) \xrightarrow{\mathbb{P}} \lambda(u_o)$ and

$$\sqrt{\mathfrak{b}T} \left(\hat{\lambda}(u_o) - \mathbb{E} \hat{\lambda}(u_o) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \lambda(u_o) \int_{\mathbb{R}} \mathfrak{K}^2(x) dx \right)$$

as $T \rightarrow \infty$, for $u_o \in (0, 1)$. Moreover, if $\mathfrak{b}^{2(m'+\gamma')+1}T = o(1)$, then

$$\sqrt{\mathfrak{b}T} \left(\mathbb{E} \hat{\lambda}(u_o) - \lambda(u_o) - \frac{\mathfrak{b}^2}{2} \lambda^{(2)}(u_o) \int_{\mathbb{R}} x^2 \mathfrak{K}(x) dx \cdot I_{\{m'=2\}} \right) = o_p(1),$$

particularly

$$\sqrt{\mathfrak{b}T} \left(\hat{\lambda}(u_o) - \lambda(u_o) - \frac{\mathfrak{b}^2}{2} \lambda^{(2)}(u_o) \int_{\mathbb{R}} x^2 \mathfrak{K}(x) dx \cdot I_{\{m'=2\}} \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \lambda(u_o) \int_{\mathbb{R}} \mathfrak{K}^2(x) dx \right).$$

As discussed in Section 2 we compare a classical pre-averaging estimator for $\sigma_{clock}^2(u_o)$ with a new estimator based on the decomposition formula (8). The classical estimator in the infill-framework is defined by

$$\begin{aligned} \hat{\sigma}_{clock,pavg}^2(u_o) &:= \frac{1}{bH} \frac{1}{g_2} \sum_{i=1}^{N_T} K \left(\frac{t_i - u_o T}{bT} \right) (\overline{\Delta Y}_{t_i,T})^2 \\ &\quad - \frac{1}{2bH} \frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2} \sum_{i=1}^{N_T} K \left(\frac{t_i - u_o T}{bT} \right) (Y_{t_i,T} - Y_{t_{i-1},T})^2, \end{aligned} \quad (10)$$

with pre-averaging steps $\overline{\Delta Y}_{t_i,T}$ given by

$$\overline{\Delta Y}_{t_i,T} := \sum_{l=1}^{H-1} g \left(\frac{l}{H} \right) (Y_{t_{i+l},T} - Y_{t_{i+l-1},T}) = - \sum_{l=1}^{H-1} h \left(\frac{l}{H} \right) Y_{t_{i+l},T}.$$

We define $h(l/H) := g((l+1)/H) - g(l/H)$, where g is another differentiable weight function defined on $[0, 1]$ with $g(0) = g(1) = 0$ and which has piecewise Lipschitz continuous derivatives $g^{(1)}$. The kernel function K and the bandwidth b are similar to those of the intensity estimate, i.e. K satisfies

condition (K) and b depends on the time span T such that $b \rightarrow 0$ and $bT \rightarrow \infty$ as $T \rightarrow \infty$. Moreover the pre-averaging block size $H = H(b)$ also depends on T such that $H \rightarrow \infty$ and $H/bT \rightarrow 0$ as $T \rightarrow \infty$. We assume that the limits $\sum_{l=1}^{H-1} g(l/H)^n/H$ and $\sum_{l=1}^{H-1} g^{(1)}(l/H)^n/H$ exist and equal $g_n := \int_0^1 g(x)^n dx$ and $g'_n := \int_0^1 g^{(1)}(x)^n dx$ respectively, for $n \in \mathbb{N}$.

The subscript *pavg* stands for the name of the procedure, *pre-averaging*. As its name suggests, we first calculate the average of log returns weighted by a function g over each block of size H and then apply a local sum of squares of these averages (similar to the filtering of realized volatility) to construct a spot volatility estimate. By doing this, the variance of the noise is reduced by a factor of $1/H$, as can be seen in our proof (see also Jacod et al. 2009). In particular, this block size H will play a crucial role in this setting along with the main bandwidth size b . Finally, a bias term induced by the additional measurement error will be corrected by the second term of (10).

3.4 Theorem Suppose that Assumption 3.1 is fulfilled. Let K have bounded first derivatives and let the block size $H = \delta \cdot T^{1/2}$ for $\delta \in (0, \infty)$ and $b^{2\alpha+1}T^{1/2} = o(1)$ with $\alpha = \min\{m + \gamma, m' + \gamma'\}$. Then

$$\sqrt{bT^{1/2}} \{\hat{\sigma}_{clock,pavg}^2(u_o) - \sigma_{clock}^2(u_o) - BIAS\} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \delta\eta_A^2 + \frac{1}{\delta}\eta_B^2 + \frac{1}{\delta^3}\eta_C^2\right)$$

as $T \rightarrow \infty$, for $u_o \in (0, 1)$, where

$$\begin{aligned} \eta_A^2 &= 2\sigma^4(u_o)\lambda(u_o) \int_{\mathbb{R}} K^2(x)dx, & \eta_B^2 &= 4\omega^2\sigma^2(u_o)\lambda(u_o)(g'_2/g_2) \int_{\mathbb{R}} K^2(x)dx, \\ \eta_C^2 &= 2\omega^4\lambda(u_o)(g'_2/g_2)^2 \int_{\mathbb{R}} K^2(x)dx, & BIAS &= \frac{1}{2} (\sigma^2(u_o)\lambda(u_o))^{(2)} b^2 \int_{\mathbb{R}} x^2 K(x)dx \cdot I_{\{m=m'=2\}}. \end{aligned}$$

We see that our result based on the transaction-time model is different from the classical result based on the standard diffusion model in the way that the asymptotic bias and variance rely on the transaction intensity. The consistency of this estimate has been implicitly proven, leading to many consistent estimators for functionals of spot volatility (including the integrated volatility) by using continuous-mapping theorem. For example, the unknown component in the asymptotic variance $\sigma^4(u_o)\lambda(u_o)$ can be estimated by the square of $\hat{\sigma}_{clock,pavg}^2(u_o)$ divided by $\hat{\lambda}(u_o)$. Alternatively one could construct another consistent estimator for the asymptotic variance by applying the pre-averaging technique (Jacod et al. 2009, eq. (3.7)).

Due to the concept of volatility decomposition, we now formulate an alternative estimator for the clock-time volatility with the following product

$$\tilde{\sigma}_{clock,pavg}^2(u_o) := \hat{\sigma}_{pavg}^2(u_o) \cdot \hat{\lambda}(u_o). \quad (11)$$

For this product, the intensity estimate (9) is used and the tick-time volatility estimator is given by

$$\begin{aligned} \hat{\sigma}_{pavg}^2(u_o) &:= \frac{T}{NH} \frac{1}{g_2} \sum_{i=i_o-N}^{i_o+N} k\left(\frac{i-i_o}{N}\right) (\overline{\Delta Y}_{t_i,T})^2 \\ &\quad - \frac{T}{2NH} \frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2} \sum_{i=i_o-N}^{i_o+N} k\left(\frac{i-i_o}{N}\right) (Y_{t_i,T} - Y_{t_{i-1},T})^2, \end{aligned} \quad (12)$$

where $i_o := \inf\{i : t_i \geq u_o T\}$, i.e. t_{i_o} is the first arrival time after or at the time point of interest t_o . The kernel function k and the segment length N satisfy Condition (K) and $N = N(T) \rightarrow \infty$ and $N/T \rightarrow 0$ as $T \rightarrow \infty$. In fact, this estimator also relies on the pre-averaging approach, and therefore

the same conditions for g and h are taken from the preceding clock-time volatility estimator with the block size $H = H(N)$ satisfying $H \rightarrow \infty$ and $H/N \rightarrow 0$ as $T \rightarrow \infty$. Note that we have used the same letters g , h , and H for both estimators in order to not complicate the notation.

At first sight, both estimates $\hat{\sigma}_{clock,pavg}^2(\cdot)$ and $\hat{\sigma}_{pavg}^2(\cdot)$ look very similar, as they are based on the filtering of pre-averaging estimators. However, there is a distinction between these two: one is based on tick time and the other is based on clock time. More precisely, in (12) the (exactly) N -nearest observed pre-averaged terms from both sides of the considered time point $u_o = t_o/T$ are taken into account so that the influence of the arrival rate is removed, while K in (10) uses all pre-averaged terms inside the interval $[t_o - bT, t_o + bT]$. As a matter of fact, the number of transactions/pre-averaged terms over this interval is random and depends on the trading intensity.

3.5 Theorem *Under Assumption 3.1, the pre-filtering block size $H = \delta \cdot T^{1/2}$ for $\delta \in (0, \infty)$, and k has bounded first derivatives, we obtain*

$$\sqrt{\frac{N}{T^{1/2}}} \{ \hat{\sigma}_{pavg}^2(u_o) - \sigma^2(u_o) \} \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \delta \xi_A^2 + \frac{1}{\delta} \xi_B^2 + \frac{1}{\delta^3} \xi_C^2 \right)$$

where

$$\xi_A^2 = 2\sigma^4(u_o) \int_{\mathbb{R}} k^2(x) dx, \quad \xi_B^2 = 4\omega^2 \sigma^2(u_o) \frac{g_2'}{g_2} \int_{\mathbb{R}} k^2(x) dx \quad \text{and} \quad \xi_C^2 = 2\omega^4 \left(\frac{g_2'}{g_2} \right)^2 \int_{\mathbb{R}} k^2(x) dx,$$

under the segment conditions

$$\begin{aligned} N^{1+2\gamma}/T^{1/2+2\gamma} &\rightarrow 0 & \text{for } m = 0, & & N^{3+\gamma^*}/T^{5/2+\gamma^*} &\rightarrow 0 & \text{for } m = 1 \text{ and } m' = 0, \\ N^{3+\gamma}/T^{5/2+\gamma} &\rightarrow 0 & \text{for } m = 1 \text{ and } m' = 1, 2, & & N^{3+\gamma'}/T^{5/2+\gamma'} &\rightarrow 0 & \text{for } m = 2 \text{ and } m' = 0, \\ N^4/T^{7/2} &\rightarrow 0 & \text{for } m = 2 \text{ and } m' = 1, 2, & & & & \end{aligned} \quad (13)$$

with $\gamma^* = \min\{\gamma, \gamma'\}$ as $T \rightarrow \infty$.

It is remarkable that the bias derivation for this estimate is much more complicated than that of $\hat{\sigma}_{clock,pavg}^2(\cdot)$ in the case of higher orders of smoothness ($\min(m, m') \geq 1$). In particular, the explicit bias term $BIAS$ in Theorem 3.4 cannot be stated, even though the kernel function is symmetric and the parameter functions are twice differentiable ($m, m' = 2$). This weakness leads to a reduction in the rate of convergence in many cases, particularly when $\min(m, m') \geq 1$ (more precisely, $(I_{4,2,1})$ and $(I_{4,2,2})$ in the proof of Theorem 3.5 only vanish if $N^4/(HT^3) \rightarrow 0$). The above segment conditions are given to enable us to neglect some asymptotic bias terms.

From (11), the alternative estimator is clearly consistent, since it is the product of two consistent estimators. The limit distribution is given below, where the rate of convergence will depend on the convergence rate bT and $N/T^{1/2}$ in Theorems 3.3 and 3.5 respectively. For example, if $N/T^{1/2} = o(bT)$, we have

$$\begin{aligned} &\sqrt{\frac{N}{T^{1/2}}} \{ \hat{\sigma}_{clock,pavg}^2(u_o) - \sigma_{clock}^2(u_o) \} \\ &= \hat{\lambda}(u_o) \sqrt{\frac{N}{T^{1/2}}} \{ \hat{\sigma}_{pavg}^2(u_o) - \sigma^2(u_o) \} + \sigma^2(u_o) \sqrt{\frac{N}{T^{1/2}}} \frac{1}{\sqrt{bT}} \cdot \sqrt{bT} \{ \hat{\lambda}(u_o) - \lambda(u_o) - BIAS_{\lambda} \} \\ &\quad + \sigma^2(u_o) \sqrt{\frac{N}{T^{1/2}}} \cdot BIAS_{\lambda} \\ &= \hat{\lambda}(u_o) \sqrt{\frac{N}{T^{1/2}}} \{ \hat{\sigma}_{pavg}^2(u_o) - \sigma^2(u_o) \} + o(1) + \sigma^2(u_o) \sqrt{\frac{N}{T^{1/2}}} \cdot BIAS_{\lambda} \end{aligned}$$

Thereby, the resulting limit distribution is dominated by the limit of $\hat{\sigma}_{pavg}^2(\cdot)$ given in the last theorem, not that of $\hat{\lambda}(\cdot)$.

3.6 Theorem *Let all assumptions be satisfied and suppose that the bandwidth \mathfrak{b} and the segment length N fulfill the conditions given in Theorem 3.3 and 3.5 respectively. For $u_o \in (0, 1)$ we obtain*

$$\sqrt{\frac{N}{T^{1/2}}} \{ \tilde{\sigma}_{clock,pavg}^2(u_o) - \sigma_{clock}^2(u_o) \} \xrightarrow{\mathcal{D}} \mathcal{N}(0, V^2)$$

with

$$V^2 = \lambda^2(u_o) \left\{ \delta \xi_A^2 + \frac{1}{\delta} \xi_B^2 + \frac{1}{\delta^3} \xi_C^2 \right\} I_{\{m=m'=0, \gamma' > \frac{\gamma}{2\gamma+2} \text{ or } m=1, m'=0, \gamma' > \frac{\gamma^*+2}{2\gamma^*+8} \text{ or } m=2, m'=0, \gamma' > \frac{\sqrt{65}-7}{4} \text{ or } m'=1, 2\}},$$

and

$$\sqrt{\mathfrak{b}T} \{ \tilde{\sigma}_{clock,pavg}^2(u_o) - \sigma_{clock}^2(u_o) \} \xrightarrow{\mathcal{D}} \mathcal{N}(0, W^2)$$

with

$$W^2 = \sigma^4(u_o) \lambda(u_o) \int_{\mathbb{R}} \mathfrak{K}^2(x) dx I_{\{m=m'=0, \gamma' \leq \frac{\gamma}{2\gamma+2} \text{ or } m=1, m'=0, \gamma' \leq \frac{\gamma^*+2}{2\gamma^*+8} \text{ or } m=2, m'=0, \gamma' \leq \frac{\sqrt{65}-7}{4}\}} \\ + c_1 \lambda^2(u_o) \left\{ \delta \xi_A^2 + \frac{1}{\delta} \xi_B^2 + \frac{1}{\delta^3} \xi_C^2 \right\} I_{\{m=m'=0, \gamma' = \frac{\gamma}{2\gamma+2} \text{ or } m=1, m'=0, \gamma' = \frac{\gamma^*+2}{2\gamma^*+8} \text{ or } m=2, m'=0, \gamma' = \frac{\sqrt{65}-7}{4}\}},$$

where $c_1 := \mathfrak{b}T/(N/T^{1/2})$ if $\mathfrak{b}T$ and $N/T^{1/2}$ are of the same order.

3.7 Remark i) For the pre-averaging steps, the block size H in (10) may differ from H in (12) (both are related to T). For this theoretical investigation, we select $H = O(T^{1/2})$ to balance the rate of convergence of the limit distributions (I), (II), and (III) (see the beginning of the proofs of Theorems 3.4 and 3.5 in section A.3) in order to obtain those asymptotic normality results.

ii) From a practical point of view, all of the unknown components in the asymptotic variances can be estimated by using the existing statistics presented in this section. The variance ω^2 of the microstructure noise can be estimated by $\hat{\sigma}_{clock}^2(u)/(2T\hat{\lambda}(u))$, where $\hat{\sigma}_{clock}^2(u) := \sum_{i=1}^{N_T} \frac{1}{b} K\left(\frac{t_i - uT}{bT}\right) (Y_{t_i, T} - Y_{t_{i-1}, T})^2$ is a filtered realized volatility, since $\hat{\sigma}_{clock}^2(u) = 2T\omega^2\lambda(u) + o_p(1)$. The data-adaptive choice of the smoothing parameters b , \mathfrak{b} , N , $H(b)$, and $H(N)$ remains to be solved.

It is well-known that the presence of microstructure noise causes a reduction in the rate of convergence of volatility estimation. The advantage of our decomposable estimator with respect to microstructure noise is, that the noise does not disturb the transaction-times but only the transaction-prices. In Section 2, our empirical analysis suggests that the tick-time volatility curve is in general less fluctuating than the intensity curve. Therefore coping with microstructure noise becomes easier, as we may choose a larger window for $\hat{\sigma}_{pavg}^2(\cdot)$ with effectively more data than with $\hat{\sigma}_{pavg, clock}^2(\cdot)$, which is as rough as the intensity curve. Mathematically, this leads to a higher rate of convergence of the volatility estimator (in some cases even better than the lower bound of the estimation for spot volatility in the standard noisy model - see below).

From this point of view, we now compare the performance of the estimates $\hat{\sigma}_{clock,pavg}^2(\cdot)$ (based on the classical pre-averaging method) and $\tilde{\sigma}_{clock,pavg}^2(\cdot)$ (based on the volatility decomposition). The same kernel functions \mathfrak{K} , k , and K and weighting function g are used in both estimates. Before discussing the comparison which is summarized in Table 1, we explicitly demonstrate one of those

Table 1: Comparison of rates of convergence and asymptotic variances between spot volatility estimators

Case	Conditions		Rate($\tilde{\sigma}_{clock,pavg}^2$) vs. Rate($\hat{\sigma}_{clock,pavg}^2$)	Var [$\tilde{\sigma}_{clock,pavg}^2$] vs. Var [$\hat{\sigma}_{clock,pavg}^2$]
c1	$m = 0, m' = 0$	$\gamma > \gamma'$	faster	
c2	$m = 1, m' = 0$	-		
c3	$m = 1, m' = 1$	$2\gamma' < \gamma$		
c4	$m = 2, m' = 0$	-		
c5	$m = 2, m' = 1$	$\gamma' < 1/2$		
c6	$m = 0, m' = 0$	$\gamma \leq \gamma'$	same	smaller, if $\lambda(\cdot) < 1$; larger, otherwise.
c7	$m = 0, m' = 1$	-		
c8	$m = 0, m' = 2$	-		
c9	$m = 1, m' = 1$	$\gamma \leq 2\gamma'$	unknown	
c10	$m = 1, m' = 2$	-		
c11	$m = 2, m' = 1$	$\gamma' \geq 1/2$		
c12	$m = 2, m' = 2$	-		

cases in detail (the other cases can be done similarly). Let $m = m' = 0$ and $\gamma' \leq \gamma/2(\gamma + 1)$. According to Theorems 3.6 and 3.4 we have

$$\sqrt{bT} \{ \tilde{\sigma}_{clock,pavg}^2(u_o) - \sigma_{clock}^2(u_o) \} \xrightarrow{\mathcal{D}} \mathcal{N}(0, W^2)$$

under the bandwidth condition $b^{2\gamma'+1}T \rightarrow 0$, and

$$\sqrt{bT^{1/2}} \{ \tilde{\sigma}_{clock,pavg}^2(u_o) - \sigma_{clock}^2(u_o) \} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \delta\eta_A^2 + \frac{1}{\delta}\eta_B^2 + \frac{1}{\delta^3}\eta_C^2)$$

under $b^{2\gamma'+1}T^{1/2} \rightarrow 0$. These constraints imply

$$bT = o(T^{\frac{2\gamma'}{2\gamma'+1}}) \quad \text{and} \quad bT^{1/2} = o(T^{\frac{\gamma'}{2\gamma'+1}}),$$

meaning that the rate of convergence of $\tilde{\sigma}_{clock,pavg}^2(\cdot)$ is much faster than that of $\hat{\sigma}_{clock,pavg}^2(\cdot)$.

Moreover, in the standard diffusion model with the presence of noise, a lower bound for spot volatility estimation is derived in a minimax sense with respect to the L_2 -loss function. This equals $n^{\frac{-\alpha}{2\alpha+1}}$, given a Hölder-exponent of α for the spot volatility function, where n is the number of subdivisions (Munk and Schmidt-Hieber 2010). Our estimate $\tilde{\sigma}_{clock,pavg}^2(\cdot)$ is better than that bound in this particular case and in many other cases, see Table 1. Thus, the approach based on volatility decomposition of transaction-time models outperforms previous approaches applied to the standard model in these cases (provided that the time-rescaled model of this paper is correct).

Table 1 compares the performance of $\tilde{\sigma}_{clock,pavg}^2(\cdot)$ and $\hat{\sigma}_{clock,pavg}^2(\cdot)$ in terms of their rates of convergence and asymptotic variances. We see that the alternative volatility estimator outperforms the standard estimator in the sense that its rate of convergence is improved, see c1–c5. In these cases, the smoothness order of the tick-time volatility is higher than that of the clock-time volatility and intensity (this is clear from above, as we can choose larger window for estimating tick-time volatility than for estimating clock-time volatility). In fact, in the empirical examples in Section 2 the tick-time volatility curve is smoother than the transaction intensity curve. Thus, we can significantly improve volatility estimation by considering the volatility decomposition. In c6–c8, both estimators

possess the same rate of convergence, but the asymptotic variance of $\tilde{\sigma}_{clock,pavg}^2(\cdot)$ turns out to be smaller if the intensity is less than 1; otherwise $\hat{\sigma}_{clock,pavg}^2(\cdot)$ yields better results. Lastly, in cases c9–c12 the situation is unknown since it is very difficult to derive the bias of $\hat{\sigma}_{pavg}^2(\cdot)$ explicitly in this situation when $\min(m, m') \geq 1$. The conditions needed to show that the bias is of lower order indicate however, that the alternative estimator has a slower rate of convergence than the classical one.

4. CONCLUSION

In this paper we have advocated the use of a spot volatility estimate based on a volatility decomposition in a time-changed price-model according to the trading times. In this model clock-time volatility splits up into the product of two curves, namely tick-time volatility and trading intensity. Both curves can be identified and we have argued that both curves contain valuable information about the original volatility curve. For example U-shape and the increase of volatility at the end of the trading day are in our opinion solely features of trading intensity while the influence of company related news mainly hits tick-time volatility or both curves.

An important finding in our view is that the tick-time volatility curve is often much smoother than the clock-time volatility curve of high-liquid equities. This means that the major part of fluctuations in clock-time volatility is due to fluctuations of the trading intensity.

There is an important consequence of these findings also for statistical inference: microstructure noise does not influence the estimate of trading intensity but only the estimator of tick-time volatility. Since this usually is the smoother curve we may choose a larger bandwidth with the benefit of a faster rate of convergence and a better coping of microstructure noise. In particular, one may outperform the rate of convergence of the optimal estimator in the classical diffusion model.

For the mathematical investigation of this model we have introduced an infill asymptotic approach, and derived the asymptotic properties of the new estimator in the case of a deterministic volatility curve and a deterministic intensity curve of a point process. If both curves are replaced by stochastic processes one may use ideas along the lines of the work by Koo and Linton (2012) on locally stationary diffusion models in combination with similar models for the point process of transaction times such as in Roueff et al. (2016). A comprehensive treatment of this situation seems challenging and beyond the scope of this paper.

From an applied point of view it is also of high interest to find proper models where the trading intensity depends on past log-prices, and where the volatility has some GARCH-type structure - possibly depending in addition on the past intensity of the point process.

APPENDIX

A.1 Details of the Data Analysis

In our empirical study we use tick-resolution trading history provided by the data vendor *Quant-Quote TickView*. We analyze intraday transaction data from the NASDAQ stock exchange.

In general we clean raw data before analyzing it in the following main steps: i) deleting all pre- and after-market data, i.e. only transactions between 09:30 AM–04:00 PM are considered; ii) filtering raw data from the outliers, such as price errors, and deleting entries with abnormal sale conditions. Although the accuracy of transaction time is down to milliseconds, the resolution of timestamps is limited to only one second, which leads to the possibility of having multiple consecutive transactions occurring at the same time. Nevertheless, the order of the trades is correctly placed. In the case of multiple trades, which is often the case for liquidly traded equities, these time points will be separated into equally-spaced times, for example t_{10} , t_{11} and t_{12} occurring at time 34210 (= 09:30:10

AM) are adjusted to $t_{10} = 34210$, $t_{11} = 34210.33$, $t_{12} = 34210.67$. We note that there are 6.5 market hours in a trading day, which is equal to $T = 23,400$ seconds.

Figures 1–4 show the analysis of transaction data on April 1, 2014; MSFT = Microsoft (25,198 transactions), GM = General Motors (31,044 transactions), HON = Honeywell (4,162 transactions) and NKE = Nike (4,341 transactions) with the estimators described in Section 2 (remember that these estimators differ from the estimators from Section 3 by the factor $1/T$). For the investigation of the high-liquid stocks in Figures 1–2 we chose the time-bandwidth $M = 200$ (the estimate uses $2M = 400$), the tick-bandwidth $m = \lfloor 200 \cdot \{\# \text{ of trades} \} / T \rfloor$ (resulting in $m = 215$ for MSFT; $m = 265$ for GM) and $H = 15$, i.e. M and m were chosen to cover the same range (for $\hat{\sigma}_{clock,pavg}^2(t_o)$, all observed pre-averaged terms $\bar{\Delta}Y_{t_i}$ inside the interval $(t_o - M, t_o + M]$ contribute to the estimator at time t_o , whereas exactly m observed pre-averaged terms from the left- and the right-hand side of t_o contribute to the estimator $\hat{\sigma}_{pavg}^2(t_o)$; therefore it does not depend on the transaction intensity). For the less liquid stocks HON and NKE in Figures 3–4 we chose $M = 300$, $m = \lfloor 300 \cdot \{\# \text{ of trades} \} / T \rfloor$ (resulting in $m = 53$ for HON; $m = 55$ for NKE) and $H = 15$. The weighting functions $\mathfrak{K}(x)$, $K(x)$, and $k(x)$ applied here are Epanechnikov kernels $\frac{3}{4}(1-x^2)I_{\{|x| \leq 1\}}$ and the weighting function $g(x)$ in the pre-averaging steps is $g(x) = x(1-x)I_{\{0 \leq x \leq 1\}}$.

The bandwidth for the green estimator in Figures 1–6 is chosen adaptively as described in Dahlhaus and Neddermeyer (2013). This (recursive) estimator is equivalent to a kernel estimator with a one-sided kernel which results in an additional bias. This bias has been corrected in the above plots by a time-shift.

For the simulation in Figures 5–6 we have used additive noise with rounding, namely

$$Y_{t_i} = \log(\lfloor 100 \cdot (\exp(X_{t_i}) + \varepsilon_i) \rfloor / 100)$$

with $\varepsilon_i \sim \mathcal{N}(0, 0.001^2)$. The tick-time volatility is chosen to be $\sigma^2(t) = \exp(-18)$ in Figure 5 and $\sigma^2(t) = \exp(-18 + \cos(10\pi t/T))$ in Figure 6 to mimic the shape of the empirical volatility per tick as in MSFT and NKE respectively. Furthermore, the transaction arrivals used here are taken from a real stock - CSCO on April 1, 2014, in order to mimic a real trading intensity.

A.2 Proof of the Volatility Decomposition

Proof of Proposition 2.2. For $\delta > 0$ and $t \in (0, T)$, we obtain with

$$t_{i_l} := \inf\{t_i : t_i > t\} \quad \text{and} \quad t_{i_u} := \sup\{t_i : t_i \leq t + \delta\}$$

$$\begin{aligned} \mathbb{E} \left[(X_{t+\delta} - X_t)^2 \mid \mathcal{F}_t \right] &= \mathbb{E} \left[\left(\sum_{t < t_i \leq t+\delta} \sigma_{t_i} U_i \right)^2 \mid \mathcal{F}_t \right] \\ &\stackrel{(*)}{=} \mathbb{E} \left[\left(\sum_{t < t_i < t_{i_u}} \sigma_{t_i} U_i \right)^2 \mid \mathcal{F}_t \right] + \mathbb{E} \left[2 \left(\sum_{t < t_i < t_{i_u}} \sigma_{t_i} U_i \right) \mathbb{E} [\sigma_{t_{i_u}} U_{i_u} \mid \mathcal{F}_{t_{i_u}-}] \mid \mathcal{F}_t \right] \\ &\quad + \mathbb{E} \left[\mathbb{E} [\sigma_{t_{i_u}}^2 U_{i_u}^2 \mid \mathcal{F}_{t_{i_u}-}] \mid \mathcal{F}_t \right] \\ &= \mathbb{E} \left[\left(\sum_{t < t_i < t_{i_u}} \sigma_{t_i} U_i \right)^2 \mid \mathcal{F}_t \right] + \mathbb{E} [\sigma_{t_{i_u}}^2 \mid \mathcal{F}_t] \\ &= \dots = \mathbb{E} [\sigma_{t_{i_u}}^2 + \sigma_{t_{i_u-1}}^2 + \dots + \sigma_{t_{i_l}}^2 \mid \mathcal{F}_t] \\ &= \mathbb{E} \left[\int_t^{t+\delta} \sigma_s^2 dN_s \mid \mathcal{F}_t \right] = \mathbb{E} \left[\int_t^{t+\delta} \sigma_s^2 \cdot \lambda_s ds \mid \mathcal{F}_t \right] \end{aligned}$$

which implies the assertion due to the dominated convergence theorem as the processes σ_t and λ_t are continuous over $[0, T]$. The last equality holds by the martingale property of $\int_0^t \sigma_s dM_s$, since $M_t := N_t - \int_0^t \lambda_s ds$ is a martingale and therefore the stochastic integral with M_t as integrator is also a martingale. In (*) we can take $\sigma_{t_{iu}}$ out of the conditional expectation, since we have assumed that σ_t is \mathcal{F}_t -predictable, so σ_{t_i} is \mathcal{F}_{t_i-} -measurable. \square

A.3 Proofs for the Rescaled Model

In what follows, some steps of the proofs are related to martingale theory. In fact, the same results can be obtained without using it, however the calculations could be long and cumbersome. Another benefit of using this theory in our framework is that we can extend the proofs more easily to a more general case of stochastic intensity models where martingale dynamics are needed. We define $M_{t,T} := N_{t,T} - \int_0^t \lambda(s/T) ds$, which is clearly a martingale. Then, it is possible to define a stochastic integral $\int_0^t c_{s,T} dM_{s,T}$, where $c_{t,T}$ is a predictable process. We can show that

$$\mathbb{E} \left[\int_0^t c_{s,T} dM_{s,T} \right]^2 = \mathbb{E} \left[\int_0^t c_{s,T} \lambda(s/T) ds \right] \quad (\text{A.1})$$

(e.g. Kuo 2006, chap. 6). Throughout the proofs, C is used as a generic constant.

Proof of Theorem 3.3. By the isometry (A.1) we have

$$\begin{aligned} \mathbb{E} \left[\hat{\lambda}(u_o) - \frac{1}{bT} \int_0^T \mathfrak{K} \left(\frac{t - u_o T}{bT} \right) \lambda \left(\frac{t}{T} \right) dt \right]^2 &= \frac{1}{b^2 T^2} \mathbb{E} \left[\int_0^T \mathfrak{K} \left(\frac{t - u_o T}{bT} \right) dM_{t,T} \right]^2 \\ &= \frac{1}{b^2 T^2} \int_0^T \mathfrak{K}^2 \left(\frac{t - u_o T}{bT} \right) \lambda \left(\frac{t}{T} \right) dt \rightarrow 0 \end{aligned}$$

for $b \leq u_o \leq 1 - b$ with $0 < b \leq 1/2$. A usual bias calculation gives

$$\frac{1}{bT} \int_0^T \mathfrak{K} \left(\frac{t - u_o T}{bT} \right) \lambda \left(\frac{t}{T} \right) dt - \lambda(u_o) = \int_{\mathbb{R}} \mathfrak{K}(x) \{ \lambda(u_o + xb) - \lambda(u_o) \} dx,$$

which implies the consistency of the intensity estimate, since $\lambda(\cdot)$ is bounded continuous. To show the asymptotic normality we use the limit theorem for triangular arrays. First we divide $[0, T]$ into M_T equidistant subintervals of a fixed length Δ , i.e. $M_T = \lfloor T/\Delta \rfloor \rightarrow \infty$, as $T \rightarrow \infty$. For $j = 0, 1, \dots, M_T$ we define $\Delta_j := j\Delta$ and rewrite

$$\hat{\lambda}(u_o) = \sum_{i=1}^{M_T} \int_{\Delta_{i-1}}^{\Delta_i} \frac{1}{bT} \mathfrak{K} \left(\frac{t - u_o T}{bT} \right) dN_{t,T} + \int_{\Delta_{M_T}}^T \frac{1}{bT} \mathfrak{K} \left(\frac{t - u_o T}{bT} \right) dN_{t,T},$$

which gives

$$\begin{aligned} \sqrt{bT} \left(\hat{\lambda}(u_o) - \frac{1}{bT} \int_0^T \mathfrak{K} \left(\frac{t - u_o T}{bT} \right) \lambda \left(\frac{t}{T} \right) dt \right) &= \sum_{i=1}^{M_T} \int_{\Delta_{i-1}}^{\Delta_i} \frac{1}{\sqrt{bT}} \mathfrak{K} \left(\frac{t - u_o T}{bT} \right) dM_{t,T} + R \\ &=: \sum_{i=1}^{M_T} Z_{i,T} + R. \end{aligned} \quad (\text{A.2})$$

Owing to independent increments of $N_{t,T}$, the random variables $Z_{i,T}$, $i = 1, \dots, M_T$, are independent with $\mathbb{E}[Z_{i,T}] = 0$. We can show that (i)

$$V_{M_T}^2 := \sum_{i=1}^{M_T} \mathbb{E}[Z_{i,T}^2] \rightarrow \lambda(u_o) \int_{\mathbb{R}} \mathfrak{K}^2(x) dx =: V^2;$$

and (ii) (Lindeberg's condition) for all $\varepsilon > 0$,

$$\frac{1}{V_{M_T}^2} \sum_{i=1}^{M_T} \mathbb{E} \left[Z_{i,T}^2 I_{\{|Z_{i,T}| \geq \varepsilon \cdot V_{M_T}\}} \right] = \frac{1}{V_{M_T}^2} \sum_{i=1}^{M_T} \mathbb{E} \left[Z_{i,T}^2 I_{\left\{ \left| \int_{\Delta_{i-1}}^{\Delta_i} \mathfrak{K} \left(\frac{t - u_o T}{bT} \right) dM_{t,T} \right| \geq \sqrt{bT} \cdot \varepsilon \cdot V_{M_T} \right\}} \right] \rightarrow 0$$

by the dominated convergence theorem, since $\mathbb{E} [Z_{i,T}^2] < \infty$. Thus $\sum_{i=1}^{M_T} Z_{i,T} \xrightarrow{\mathcal{D}} \mathcal{N}(0, V^2)$ which implies that (A.2) $\xrightarrow{\mathcal{D}} \mathcal{N}(0, V^2)$, since the rest term R is asymptotically negligible. Similarly,

$$\mathbb{E} [\widehat{\lambda}(u_o)] - \lambda(u_o) = \frac{1}{2} \lambda^{(2)}(u_o) \int_{\mathbb{R}} b^2 x^2 \mathfrak{K}(x) dx \cdot I_{\{m'=2\}} + O \left(\int_{\mathbb{R}} \mathfrak{K}(x) |x b|^{m'+\gamma'} dx \right),$$

since $\lambda(\cdot)$ lies in $\mathcal{C}^{m',\gamma'}$. The assertion is then verified by the condition $b^{2(m'+\gamma')+1}T = o(1)$. \square

Proof of Theorem 3.4. First we will focus on the asymptotic normality of the statistic

$$\begin{aligned} \widehat{\sigma}_{clock,pre}^2(u_o) &:= \frac{1}{b} \frac{1}{g_2} \sum_{i=0}^{M_T-1} K \left(\frac{t_{iH} - u_o T}{bT} \right) (\overline{\Delta Y}_{t_{iH},T})^2 \\ &\quad - \frac{1}{2bH} \frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2} \sum_{i=1}^{N_T} K \left(\frac{t_i - u_o T}{bT} \right) (Y_{t_i,T} - Y_{t_{i-1},T})^2 \end{aligned}$$

with $M_T := \lfloor N_T/H \rfloor$ being the random number of blocks. The first term consists of non-overlapping blocks of data so that, conditionally on $N_{\cdot,T}$, a central limit theorem for independent triangular arrays can be applied. The second term of $\widehat{\sigma}_{clock,pre}^2(\cdot)$ remains the same as that of $\widehat{\sigma}_{clock,pavg}^2(\cdot)$ and plays no role in the limit distribution (it corrects the bias caused by the additive microstructure noise). In the end of the proof, we will show that the distinction between the two estimators $\widehat{\sigma}_{clock,pre}^2(\cdot)$ and $\widehat{\sigma}_{clock,pavg}^2(\cdot)$ is asymptotically negligible so that both have the same limit distribution.

We divide the proof into the following parts. Parts (I) – (III) concern the limit distribution, where the rate of convergence will be balanced by the choice of the block size H . Parts (IV) – (V) are related to biases; especially (V) deals with the bias caused by the additional noise.

$$\begin{aligned} (I) \quad & \sqrt{\frac{bT}{H}} \left(\frac{1}{b} \frac{1}{g_2} \sum_{i=0}^{M_T-1} K \left(\frac{t_{iH} - u_o T}{bT} \right) \left((\overline{\Delta X}_{t_{iH},T})^2 - \mathbb{E} [(\overline{\Delta X}_{t_{iH},T})^2 \mid N_{\cdot,T}] \right) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \eta_A^2), \\ (II) \quad & \sqrt{bH} \left(\frac{1}{b} \frac{2}{g_2} \sum_{i=0}^{M_T-1} K \left(\frac{t_{iH} - u_o T}{bT} \right) (\overline{\Delta X}_{t_{iH},T}) (\overline{\Delta \varepsilon}_{iH}) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \eta_B^2), \quad \text{and} \\ (III) \quad & \sqrt{\frac{bH^3}{T}} \left(\frac{1}{b} \frac{1}{g_2} \sum_{i=0}^{M_T-1} K \left(\frac{t_{iH} - u_o T}{bT} \right) \left((\overline{\Delta \varepsilon}_{iH})^2 - \mathbb{E} [(\overline{\Delta \varepsilon}_{iH})^2] \right) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \eta_C^2). \end{aligned}$$

Furthermore, under the condition $H = \delta \cdot T^{1/2}$ for $\delta \in (0, \infty)$ and the condition $b^{2\alpha+1}T^{1/2} \rightarrow 0$ with $\alpha = \min \{m + \gamma, m' + \gamma'\}$, the biases are negligible in the limit, i.e.

$$\begin{aligned} (IV) \quad & \sqrt{bT^{1/2}} \left(\frac{1}{b} \frac{1}{g_2} \sum_{i=0}^{M_T-1} K \left(\frac{t_{iH} - u_o T}{bT} \right) \mathbb{E} [(\overline{\Delta X}_{t_{iH},T})^2 \mid N_{\cdot,T}] - \sigma^2(u_o) \lambda(u_o) - BIAS \right) = o_p(1), \\ (V) \quad & \sqrt{bT^{1/2}} \left(\frac{1}{b} \frac{1}{g_2} \sum_{i=0}^{M_T-1} K \left(\frac{t_{iH} - u_o T}{bT} \right) \mathbb{E} [(\overline{\Delta \varepsilon}_{iH})^2] \right. \\ & \quad \left. - \frac{1}{2bH} \frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2} \sum_{i=1}^{N_T} K \left(\frac{t_i - u_o T}{bT} \right) (Y_{t_i,T} - Y_{t_{i-1},T})^2 \right) = o_p(1). \end{aligned}$$

We set the left-hand side of (I) to be $\sum_{i=0}^{M_T-1} A_{i,T}$ where

$$A_{i,T} = \sqrt{\frac{bT}{H}} \frac{1}{b} \frac{1}{g_2} K \left(\frac{t_{iH} - u_o T}{bT} \right) \left((\overline{\Delta X}_{t_{iH},T})^2 - \mathbb{E} [(\overline{\Delta X}_{t_{iH},T})^2 \mid N_{\cdot,T}] \right).$$

We see that conditional on $N_{\cdot,T}$, $\{A_{i,T}\}_{i=0,\dots,M_T-1}$ is an independent sequence, so it is sufficient to show that (Hall and Heyde 1980, cor. 3.1)

$$(a) \sum_{i=0}^{M_T-1} \mathbb{E} \left[A_{i,T}^2 \mid N_{\cdot,T} \right] \xrightarrow{\mathbb{P}} \eta_A^2 \text{ and } (b) \sum_{i=0}^{M_T-1} \mathbb{E} \left[A_{i,T}^4 \mid N_{\cdot,T} \right] \xrightarrow{\mathbb{P}} 0.$$

Corresponding to independent increments of $X_{\cdot,T}$, it gives

$$\begin{aligned} & \sum_{i=0}^{M_T-1} \mathbb{E} \left[A_{i,T}^2 \mid N_{\cdot,T} \right] \\ &= \sum_{i=0}^{M_T-1} \frac{T}{bH} \frac{1}{g_2^2} K^2 \left(\frac{t_{iH} - u_o T}{bT} \right) \left\{ \mathbb{E} \left[(\overline{\Delta X}_{t_{iH},T})^4 \mid N_{\cdot,T} \right] - \left(\mathbb{E} \left[(\overline{\Delta X}_{t_{iH},T})^2 \mid N_{\cdot,T} \right] \right)^2 \right\} \\ &= \sum_{i=0}^{M_T-1} \frac{2}{H} \frac{1}{bT} \frac{1}{g_2^2} K^2 \left(\frac{t_{iH} - u_o T}{bT} \right) \left\{ \sum_{l=1}^{H-1} g^4 \left(\frac{l}{H} \right) \sigma^4 \left(\frac{t_{iH+l}}{T} \right) + \left(\sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \sigma^2 \left(\frac{t_{iH+l}}{T} \right) \right)^2 \right\} \\ & \hspace{15em} \text{(see (A.3) and (A.4) below)} \\ &= \sum_{i=1}^{N_T} \frac{2}{H^2} \frac{1}{bT} \frac{1}{g_2^2} K^2 \left(\frac{t_i - u_o T}{bT} \right) \sigma^4 \left(\frac{t_i}{T} \right) \left\{ \sum_{l=1}^{H-1} g^4 \left(\frac{l}{H} \right) + \left(\sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \right)^2 \right\} + o_p(1) \\ & \hspace{15em} \text{(see (A.5) below)} \\ &= \frac{2}{H^2} \frac{(\sum g^2(l/H))^2}{g_2^2} \sum_{i=1}^{N_T} \frac{1}{bT} K^2 \left(\frac{t_i - u_o T}{bT} \right) \sigma^4 \left(\frac{t_i}{T} \right) + o_p(1) \\ & \xrightarrow{\mathbb{P}} 2\sigma^4(u_o)\lambda(u_o) \int_{\mathbb{R}} K^2(x) dx. \end{aligned}$$

By independent increments of $X_{\cdot,T}$,

$$\mathbb{E} \left[(\overline{\Delta X}_{t_{iH},T})^4 \mid N_{\cdot,T} \right] = 2 \sum_{l=1}^{H-1} g^4 \left(\frac{l}{H} \right) \sigma^4 \left(\frac{t_{iH+l}}{T} \right) \frac{1}{T^2} + 3 \left(\sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \sigma^2 \left(\frac{t_{iH+l}}{T} \right) \frac{1}{T} \right)^2; \quad (\text{A.3})$$

$$\left(\mathbb{E} \left[(\overline{\Delta X}_{t_{iH},T})^2 \mid N_{\cdot,T} \right] \right)^2 = \left(\sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \sigma^2 \left(\frac{t_{iH+l}}{T} \right) \frac{1}{T} \right)^2. \quad (\text{A.4})$$

As K has bounded first derivatives and $\sigma(\cdot) \in \mathcal{C}^{m,\gamma}$,

$$\begin{aligned} & H \cdot \sum_{i=0}^{M_T-1} \frac{1}{H} \frac{1}{bT} \frac{1}{g_2^2} K^2 \left(\frac{t_{iH} - u_o T}{bT} \right) \sum_{l=1}^{H-1} g^4 \left(\frac{l}{H} \right) \sigma^4 \left(\frac{t_{iH+l}}{T} \right) \\ & - \sum_{j=1}^{N_T} \frac{1}{H} \frac{1}{bT} \frac{1}{g_2^2} K^2 \left(\frac{t_j - u_o T}{bT} \right) \sum_{l=1}^{H-1} g^4 \left(\frac{l}{H} \right) \sigma^4 \left(\frac{t_{\lfloor j/H \rfloor \cdot H + l}}{T} \right) \\ & + \sum_{j=1}^{N_T} \frac{1}{H} \frac{1}{bT} \frac{1}{g_2^2} K^2 \left(\frac{t_j - u_o T}{bT} \right) \sum_{l=1}^{H-1} g^4 \left(\frac{l}{H} \right) \sigma^4 \left(\frac{t_{\lfloor j/H \rfloor \cdot H + l}}{T} \right) \\ & - \sum_{j=1}^{N_T} \frac{1}{H} \frac{1}{bT} \frac{1}{g_2^2} K^2 \left(\frac{t_j - u_o T}{bT} \right) \sum_{l=1}^{H-1} g^4 \left(\frac{l}{H} \right) \sigma^4 \left(\frac{t_j}{T} \right) \\ & = O \left(\frac{H}{bT} \right) + O_p \left(\left| \frac{H}{T} \right|^\gamma \right) = o_p(1), \end{aligned} \quad (\text{A.5})$$

since $H/bT \rightarrow 0$. We now turn to condition (b):

$$\begin{aligned} & \sum_{i=0}^{M_T-1} \mathbb{E} \left[A_{i,T}^4 \mid N_{\cdot,T} \right] \\ & \leq C \cdot \frac{T^2}{b^2 H^2} \sum_{i=0}^{M_T-1} K^4 \left(\frac{t_{iH} - u_o T}{bT} \right) \mathbb{E} \left[\left\{ (\overline{\Delta X}_{t_{iH},T})^2 - \mathbb{E} \left[(\overline{\Delta X}_{t_{iH},T})^2 \mid N_{\cdot,T} \right] \right\}^4 \mid N_{\cdot,T} \right] \\ & \leq C \cdot \frac{T^2}{b^2 H^2} \sum_{i=0}^{M_T-1} K^4 \left(\frac{t_{iH} - u_o T}{bT} \right) \left(\sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \right)^4 \frac{1}{T^4} = o_p(1), \end{aligned}$$

since $H/bT \rightarrow 0$ and $\mathbb{E} \left[(\overline{\Delta X}_{t_{iH},T})^8 \mid N_{\cdot,T} \right] \leq C \cdot \left\{ \sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \sigma^2 \left(\frac{t_{iH}+l}{T} \right) \right\}^4 \frac{1}{T^4}$ holds. In order to derive the limit distribution in (II) we set

$$\sqrt{bH} \left(\frac{1}{b} \frac{2}{g_2} \sum_{i=0}^{M_T-1} K \left(\frac{t_{iH} - u_o T}{bT} \right) (\overline{\Delta X}_{t_{iH},T}) (\overline{\Delta \varepsilon}_{iH}) \right) =: \sum_{i=0}^{M_T-1} B_{i,T}.$$

Likewise we show that (conditional on $N_{\cdot,T}$)

$$(c) \sum_{i=0}^{M_T-1} \mathbb{E} \left[B_{i,T}^2 \mid N_{\cdot,T} \right] \xrightarrow{\mathbb{P}} \eta_B^2 \quad \text{and} \quad (d) \sum_{i=0}^{M_T-1} \mathbb{E} \left[B_{i,T}^4 \mid N_{\cdot,T} \right] \xrightarrow{\mathbb{P}} 0.$$

Since

$$\mathbb{E} \left[(\overline{\Delta \varepsilon}_{iH})^2 \right] = \mathbb{E} \left[\left(\sum_{l=1}^{H-1} h \left(\frac{l}{H} \right) \varepsilon_{iH+l} \right)^2 \right] = \omega^2 \sum_{l=1}^{H-1} h^2 \left(\frac{l}{H} \right),$$

it implies that (see (A.4))

$$\begin{aligned} \mathbb{E} \left[B_{i,T}^2 \mid N_{\cdot,T} \right] &= \frac{H}{b} \frac{4}{g_2^2} K^2 \left(\frac{t_{iH} - u_o T}{bT} \right) \mathbb{E} \left[(\overline{\Delta X}_{t_{iH},T})^2 \mid N_{\cdot,T} \right] \cdot \mathbb{E} \left[(\overline{\Delta \varepsilon}_{iH})^2 \mid N_{\cdot,T} \right] \\ &= \frac{H}{b} \frac{4}{g_2^2} K^2 \left(\frac{t_{iH} - u_o T}{bT} \right) \left(\sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \sigma^2 \left(\frac{t_{iH}+l}{T} \right) \frac{1}{T} \right) \left(\omega^2 \sum_{l=1}^{H-1} h^2 \left(\frac{l}{H} \right) \right). \end{aligned}$$

Hence (similar to (A.5))

$$\sum_{i=0}^{M_T-1} \mathbb{E} \left[B_{i,T}^2 \mid N_{\cdot,T} \right] \xrightarrow{\mathbb{P}} 4\omega^2 \sigma^2(u_o) \lambda(u_o) \frac{g_2'}{g_2} \int_{\mathbb{R}} K^2(x) dx,$$

since $\sum_{l=1}^{H-1} h^2 \left(\frac{l}{H} \right) = \sum_{l=1}^{H-1} \left\{ g^{(1)} \left(\frac{l}{H} \right) \frac{1}{H} + o \left(\frac{1}{H} \right) \right\}^2$. Similarly,

$$\begin{aligned} \sum_{i=0}^{M_T-1} \mathbb{E} \left[B_{i,T}^4 \mid N_{\cdot,T} \right] &= \sum_{i=0}^{M_T-1} \frac{H^2}{b^2} \frac{16}{g_2^4} K^4 \left(\frac{t_{iH} - u_o T}{bT} \right) \mathbb{E} \left[(\overline{\Delta X}_{t_{iH},T})^4 \mid N_{\cdot,T} \right] \cdot \mathbb{E} \left[(\overline{\Delta \varepsilon}_{iH})^4 \mid N_{\cdot,T} \right] \\ &\leq C \cdot \sum_{i=0}^{M_T-1} \frac{H^2}{b^2} K^4 \left(\frac{t_{iH} - u_o T}{bT} \right) \left(\sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \frac{1}{T} \right)^2 \left(\sum_{l=1}^{H-1} h^2 \left(\frac{l}{H} \right) \right)^2 \xrightarrow{\mathbb{P}} 0. \end{aligned}$$

We proceed analogously to show (III). We denote its left-hand side by $\sum_{i=0}^{M_T-1} C_{i,T}$ and show that

$$(e) \sum_{i=0}^{M_T-1} \mathbb{E} \left[C_{i,T}^2 \mid N_{\cdot,T} \right] \xrightarrow{\mathbb{P}} \eta_C^2 \quad \text{and} \quad (f) \sum_{i=0}^{M_T-1} \mathbb{E} \left[C_{i,T}^4 \mid N_{\cdot,T} \right] \xrightarrow{\mathbb{P}} 0.$$

On account of the assumption $\mathbb{E}[\varepsilon_i^4] = \theta\omega^4$, $\theta \in \mathbb{R}^+$, we have

$$\mathbb{E}[(\overline{\Delta\varepsilon}_{iH})^4] = (\theta - 1) \sum_{l=1}^{H-1} h^4 \left(\frac{l}{H} \right) \omega^4 + 3 \left(\sum_{l=1}^{H-1} h^2 \left(\frac{l}{H} \right) \right)^2 \omega^4,$$

thus

$$\begin{aligned} \sum_{i=0}^{M_T-1} \mathbb{E}[C_{i,T}^2 \mid N_{\cdot,T}] &= \sum_{i=0}^{M_T-1} \frac{H^3}{bT} \frac{\omega^4}{g_2^2} K^2 \left(\frac{t_{iH} - u_o T}{bT} \right) \left\{ (\theta - 1) \sum_{l=1}^{H-1} h^4 \left(\frac{l}{H} \right) + 2 \left(\sum_{l=1}^{H-1} h^2 \left(\frac{l}{H} \right) \right)^2 \right\} \\ &= \frac{1}{H} \cdot H \sum_{i=0}^{M_T-1} \frac{1}{bT} K^2 \left(\frac{t_{iH} - u_o T}{bT} \right) \frac{(\theta - 1)\omega^4}{g_2^2} \left\{ \sum_{l=1}^{H-1} \left(g' \left(\frac{l}{H} \right) \right)^4 \frac{1}{H} + o \left(\frac{1}{H} \right) \right\} \\ &\quad + \frac{2\omega^4}{g_2^2} \cdot H \sum_{i=0}^{M_T-1} \frac{1}{bT} K^2 \left(\frac{t_{iH} - u_o T}{bT} \right) \left\{ \sum_{l=1}^{H-1} \left(g' \left(\frac{l}{H} \right) \right)^2 \frac{1}{H} + o \left(\frac{1}{H} \right) \right\}^2 \\ &\xrightarrow{\mathbb{P}} 2\omega^4 \lambda(u_o) (g_2'/g_2)^2 \int_{\mathbb{R}} K^2(x) dx, \end{aligned}$$

which leads to (e). In fact, (f) can be established in the same manner as before, therefore omitted. To build the joint distribution of (I), (II) and (III) it suffices to show that

$$\sum_{i=0}^{M_T-1} a\sqrt{\delta}A_{i,T} + b\frac{1}{\sqrt{\delta}}B_{i,T} + c\frac{1}{\sqrt{\delta^3}}C_{i,T} \xrightarrow{\mathcal{D}} a\sqrt{\delta}A + b\frac{1}{\sqrt{\delta}}B + c\frac{1}{\sqrt{\delta^3}}C$$

by Cramer-Wold's theorem, for all $a, b, c \in \mathbb{R}$, where A, B and C are the limits of (I), (II) and (III) respectively. Indeed, this joint limit is a direct consequence of a) – f) where now the pre-averaging block size H is chosen to equal $\delta \cdot T^{1/2}$. To sum up, we have shown the first main part of the limit distribution, i.e.

$$\begin{aligned} &\sqrt{bT^{1/2}} \left(\sum_{i=0}^{M_T-1} \frac{1}{b} \frac{1}{g_2} K \left(\frac{t_{iH} - u_o T}{bT} \right) \left\{ (\overline{\Delta Y}_{t_{iH},T})^2 - \mathbb{E}[(\overline{\Delta X}_{t_{iH},T})^2 \mid N_{\cdot,T}] - \mathbb{E}[(\overline{\Delta\varepsilon}_{iH})^2] \right\} \right) \\ &\xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \delta\eta_A^2 + \frac{1}{\delta}\eta_B^2 + \frac{1}{\delta^3}\eta_C^2 \right), \end{aligned}$$

since $(\overline{\Delta Y}_{t_{iH},T}) = (\overline{\Delta X}_{t_{iH},T}) + (\overline{\Delta\varepsilon}_{iH,T})$.

We carry out the proof by showing the asymptotic biases in (IV) and (V). In order to derive (IV) we

see that

$$\begin{aligned}
& \sqrt{bT^{1/2}} \left(\frac{1}{b} \frac{1}{g_2} \sum_{i=0}^{M_T-1} K \left(\frac{t_{iH} - u_o T}{bT} \right) \mathbb{E} \left[(\overline{\Delta X}_{t_{iH}, T})^2 \mid N_{\cdot, T} \right] - \sigma^2(u_o) \lambda(u_o) - BIAS \right) \\
&= \sqrt{bT^{1/2}} \frac{1}{H} \frac{1}{g_2} \cdot \left(H \sum_{i=0}^{M_T-1} \frac{1}{bT} K \left(\frac{t_{iH} - u_o T}{bT} \right) \sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \sigma^2 \left(\frac{t_{iH+l}}{T} \right) \right. \\
&\quad \left. - \sum_{j=1}^{N_T} \frac{1}{bT} K \left(\frac{t_j - u_o T}{bT} \right) \sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \sigma^2 \left(\frac{t_{\lfloor j/H \rfloor \cdot H + l}}{T} \right) \right) \\
&\quad + \sqrt{bT^{1/2}} \frac{1}{H} \frac{1}{g_2} \sum_{j=1}^{N_T} \frac{1}{bT} K \left(\frac{t_j - u_o T}{bT} \right) \left(\sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \sigma^2 \left(\frac{t_{\lfloor j/H \rfloor \cdot H + l}}{T} \right) - \sigma^2 \left(\frac{t_j}{T} \right) \sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \right) \\
&\quad + \sqrt{bT^{1/2}} \left(\frac{1}{H} \frac{1}{g_2} \sum_{j=1}^{N_T} \frac{1}{bT} K \left(\frac{t_j - u_o T}{bT} \right) \sigma^2 \left(\frac{t_j}{T} \right) \sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \right. \\
&\quad \left. - \sigma^2(u_o) \lambda(u_o) - \frac{1}{2} (\sigma^2(u_o) \lambda(u_o))^{(2)} b^2 \int_{\mathbb{R}} x^2 K(x) dx I_{\{m=m'=2\}} \right) \\
&= O \left(\sqrt{bT^{1/2}} \cdot \frac{H}{bT} \right) + O_p \left(\sqrt{bT^{1/2}} \left| \frac{H}{T} \right|^\gamma \right) + (iv) = o_p(1), \tag{A.6}
\end{aligned}$$

since K has first bounded derivatives and $\sigma(\cdot) \in \mathcal{C}^{m, \gamma}$ (see also (A.5)) and

$$\begin{aligned}
(iv) &= \sqrt{bT^{1/2}} \left(\frac{1}{H} \frac{1}{g_2} \sum_{j=1}^{N_T} \frac{1}{bT} K \left(\frac{t_j - u_o T}{bT} \right) \sigma^2 \left(\frac{t_j}{T} \right) \sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \right. \\
&\quad \left. - \mathbb{E} \left[\frac{1}{H} \frac{1}{g_2} \sum_{j=1}^{N_T} \frac{1}{bT} K \left(\frac{t_j - u_o T}{bT} \right) \sigma^2 \left(\frac{t_j}{T} \right) \sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \right] \right) \\
&\quad + \sqrt{bT^{1/2}} \left(\mathbb{E} \left[\frac{1}{H} \frac{1}{g_2} \sum_{j=1}^{N_T} \frac{1}{bT} K \left(\frac{t_j - u_o T}{bT} \right) \sigma^2 \left(\frac{t_j}{T} \right) \sum_{l=1}^{H-1} g^2 \left(\frac{l}{H} \right) \right] \right. \\
&\quad \left. - \left\{ \sigma^2(u_o) \lambda(u_o) + \frac{1}{2} (\sigma^2(u_o) \lambda(u_o))^{(2)} b^2 \int_{\mathbb{R}} x^2 K(x) dx \cdot I_{\{m=m'=2\}} \right\} \right) \\
&= \sqrt{bT^{1/2}} \left(\int_0^T \frac{1}{bT} K \left(\frac{t - u_o T}{bT} \right) \sigma^2 \left(\frac{t}{T} \right) dM_{t, T} \right) \frac{\frac{1}{H} \sum g^2(l/H)}{g_2} + O \left(\sqrt{bT^{1/2}} b^{\min(\gamma, \gamma')} \right) = o_p(1).
\end{aligned}$$

The last equality in (A.6) is satisfied by the bandwidth condition $b^{2\alpha+1} T^{1/2} \rightarrow 0$. Finally, we separate (V) into two summands

$$\begin{aligned}
& \sqrt{bT^{1/2}} \left(\frac{1}{b} \frac{1}{g_2} \sum_{i=0}^{M_T-1} K \left(\frac{t_{iH} - u_o T}{bT} \right) \mathbb{E} \left[(\overline{\Delta \varepsilon}_{iH})^2 \right] - \frac{T}{H} \frac{\lambda(u_o)}{g_2} \sum_{l=1}^{H-1} h^2 \left(\frac{l}{H} \right) \omega^2 \right) \\
&+ \sqrt{bT^{1/2}} \left(\frac{T}{H} \frac{\lambda(u_o)}{g_2} \sum_{l=1}^{H-1} h^2 \left(\frac{l}{H} \right) \omega^2 - \frac{T}{2H} \frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2} \cdot \frac{1}{bT} \sum_{l=1}^{N_T} K \left(\frac{t_i - u_o T}{bT} \right) (Y_{t_i, T} - Y_{t_{i-1}, T})^2 \right) \\
&=: (v_1) + (v_2).
\end{aligned}$$

Since $\mathbb{E} \left[(\overline{\Delta \varepsilon_{iH}})^2 \right] = \sum_{l=1}^{H-1} h^2 \left(\frac{l}{H} \right) \omega^2$, direct calculations yield

$$\begin{aligned}
\mathbb{E} [(v_1)^2] &\leq C \cdot bT^{1/2} \left(\sum_{l=1}^{H-1} h^2 \left(\frac{l}{H} \right) \right)^2 \mathbb{E} \left[\frac{1}{b} \sum_{i=0}^{M_T-1} K \left(\frac{t_{iH} - u_o T}{bT} \right) - \frac{T}{H} \lambda(u_o) \right]^2 \\
&\leq C \cdot bT^{1/2} \frac{T^2}{H^2} \left(\sum_{l=1}^{H-1} h^2 \left(\frac{l}{H} \right) \right)^2 \left(\mathbb{E} \left[\frac{H}{bT} \sum_{i=0}^{M_T-1} K \left(\frac{t_{iH} - u_o T}{bT} \right) \right] - \mathbb{E} \left[\frac{H}{bT} \sum_{i=0}^{M_T-1} K \left(\frac{t_{iH} - u_o T}{bT} \right) \right] \right)^2 \\
&\quad + \left(\mathbb{E} \left[\frac{H}{bT} \sum_{i=0}^{M_T-1} K \left(\frac{t_{iH} - u_o T}{bT} \right) \right] - \lambda(u_o) \right)^2 \\
&= O \left(bT^{1/2} \frac{T^2}{H^2} \frac{1}{H^2} \frac{1}{bT} \right) + O \left(bT^{1/2} \frac{T^2}{H^2} \frac{1}{H^2} b^{2\gamma'} \right) = o(1).
\end{aligned}$$

The last term (v_2) is divided again into the sum of $(v_{2,1})$ and $(v_{2,2})$, where

$$(v_{2,1}) := \sqrt{bT^{1/2}} \left[\frac{T}{H} \frac{\lambda(u_o)}{g_2} \sum_{l=1}^{H-1} h^2 \left(\frac{l}{H} \right) \omega^2 - \frac{T}{2H} \frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2} \frac{1}{bT} \sum_{l=1}^{N_T} K \left(\frac{t_i - u_o T}{bT} \right) (\varepsilon_i - \varepsilon_{i-1})^2 \right]$$

and

$$\begin{aligned}
(v_{2,2}) &:= -\sqrt{bT^{1/2}} \left[\frac{T}{2H} \frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2} \frac{1}{bT} \sum_{l=1}^{N_T} K \left(\frac{t_i - u_o T}{bT} \right) \times \right. \\
&\quad \left. \times \left((X_{t_i, T} - X_{t_{i-1}, T})^2 + 2(X_{t_i, T} - X_{t_{i-1}, T})(\varepsilon_i - \varepsilon_{i-1}) \right) \right].
\end{aligned}$$

It is easy to show that $\mathbb{E} [(v_{2,1})^2] = o(1)$ and $\mathbb{E} [|v_{2,2}|] = o(1)$ hold, therefore $(V) = o_p(1)$. In summary, we have shown the asymptotic normality for $\hat{\sigma}_{clock, pre}^2(u_o)$.

To complete the proof, the difference between $\hat{\sigma}_{clock, avg}^2(u_o)$ and $\hat{\sigma}_{clock, pre}^2(u_o)$ needs to be determined; in particular we show that

$$\sqrt{bT^{1/2}} \{ \hat{\sigma}_{clock, pre}^2(u_o) - \hat{\sigma}_{clock, avg}^2(u_o) \} = o_p(1).$$

Since the derivative of K is bounded, it is enough to verify that (see also (A.5))

$$\begin{aligned}
&\sqrt{bT^{1/2}} \frac{1}{bH} \frac{1}{g_2} \sum_{i=1}^{N_T} K \left(\frac{t_i - u_o T}{bT} \right) \left\{ \left(\overline{\Delta Y}_{t_{\lfloor i/H \rfloor \cdot H}, T} \right)^2 - \left(\overline{\Delta Y}_{t_i, T} \right)^2 \right\} \\
&= \sqrt{bT^{1/2}} \frac{1}{bH} \frac{1}{g_2} \sum_{i=1}^{N_T} K \left(\frac{t_i - u_o T}{bT} \right) \left(\left\{ \left(\overline{\Delta X}_{t_{\lfloor i/H \rfloor \cdot H}, T} \right)^2 - \left(\overline{\Delta X}_{t_i, T} \right)^2 \right\} + \left\{ \left(\overline{\Delta \varepsilon}_{\lfloor i/H \rfloor \cdot H} \right)^2 - \left(\overline{\Delta \varepsilon}_i \right)^2 \right\} \right. \\
&\quad \left. + 2 \left\{ \left(\overline{\Delta X}_{t_{\lfloor i/H \rfloor \cdot H}, T} \right) \left(\overline{\Delta \varepsilon}_{\lfloor i/H \rfloor \cdot H} \right) - \left(\overline{\Delta X}_{t_i, T} \right) \left(\overline{\Delta \varepsilon}_i \right) \right\} \right) \\
&=: (T_1) + (T_2) + (T_3) = o_p(1).
\end{aligned} \tag{A.7}$$

We perform only the proof of the first term (T_1) . We will show below that $\mathbb{E}[(T_1)^2 \mid N_{\cdot,T}] = o_p(1)$.

$$\begin{aligned} \mathbb{E}[(T_1)^2 \mid N_{\cdot,T}] &= \frac{T^{1/2}}{bH^2} \frac{1}{g_2^2} \sum_{i,j=1}^{N_T} K\left(\frac{t_i - u_o T}{bT}\right) K\left(\frac{t_j - u_o T}{bT}\right) \times \\ &\quad \times \left(\mathbb{E} \left[\left(\overline{\Delta X}_{t_{\lfloor i/H \rfloor \cdot H, T}} \right)^2 \left(\overline{\Delta X}_{t_{\lfloor j/H \rfloor \cdot H, T}} \right)^2 \mid N_{\cdot,T} \right] - \mathbb{E} \left[\left(\overline{\Delta X}_{t_{\lfloor i/H \rfloor \cdot H, T}} \right)^2 \left(\overline{\Delta X}_{t_j, T} \right)^2 \mid N_{\cdot,T} \right] \right. \\ &\quad \left. - \mathbb{E} \left[\left(\overline{\Delta X}_{t_i, T} \right)^2 \left(\overline{\Delta X}_{t_{\lfloor j/H \rfloor \cdot H, T}} \right)^2 \mid N_{\cdot,T} \right] + \mathbb{E} \left[\left(\overline{\Delta X}_{t_i, T} \right)^2 \left(\overline{\Delta X}_{t_j, T} \right)^2 \mid N_{\cdot,T} \right] \right) \\ &=: (\Delta_1) - (\Delta_2) - (\Delta_3) + (\Delta_4). \end{aligned}$$

The first term is split up into three small terms:

$$\begin{aligned} (\Delta_1) &= \frac{T^{1/2}}{bH^2} \frac{1}{g_2^2} \sum_{i=j} K^2 \left(\frac{t_i - u_o T}{bT} \right) \mathbb{E} \left[\left(\overline{\Delta X}_{t_{\lfloor i/H \rfloor \cdot H, T}} \right)^2 \mid N_{\cdot,T} \right] \\ &\quad + \frac{T^{1/2}}{bH^2} \frac{1}{g_2^2} \sum_{|i-j| \geq H} K \left(\frac{t_i - u_o T}{bT} \right) K \left(\frac{t_j - u_o T}{bT} \right) \mathbb{E} \left[\left(\overline{\Delta X}_{t_{\lfloor i/H \rfloor \cdot H, T}} \right)^2 \mid N_{\cdot,T} \right] \cdot \mathbb{E} \left[\left(\overline{\Delta X}_{t_{\lfloor j/H \rfloor \cdot H, T}} \right)^2 \mid N_{\cdot,T} \right] \\ &\quad \quad \quad \text{(since both terms are independent.)} \\ &\quad + \frac{T^{1/2}}{bH^2} \frac{1}{g_2^2} \sum_{0 < |i-j| < H} K \left(\frac{t_i - u_o T}{bT} \right) K \left(\frac{t_j - u_o T}{bT} \right) \mathbb{E} \left[\left(\overline{\Delta X}_{t_{\lfloor i/H \rfloor \cdot H, T}} \right)^2 \left(\overline{\Delta X}_{t_{\lfloor j/H \rfloor \cdot H, T}} \right)^2 \mid N_{\cdot,T} \right] \\ &=: (\Delta_{1,1}) + (\Delta_{1,2}) + (\Delta_{1,3}). \end{aligned}$$

Likewise, we expand $(\Delta_2) = (\Delta_{2,1}) + (\Delta_{2,2}) + (\Delta_{2,3})$. It is clear that $(\Delta_{1,1})$ and $(\Delta_{2,1})$ are of smaller order, therefore neglected. Since $(\Delta_{1,2})$ is equal to $(\Delta_{2,2})$, they cancel out. Lastly,

$$\begin{aligned} &(\Delta_{1,3}) - (\Delta_{2,3}) \\ &= \frac{T^{1/2}}{bH^2} \frac{1}{g_2^2} \sum_{0 < |i-j| < H} K \left(\frac{t_i - u_o T}{bT} \right) K \left(\frac{t_j - u_o T}{bT} \right) \times \\ &\quad \times \mathbb{E} \left[\left(\overline{\Delta X}_{t_{\lfloor i/H \rfloor \cdot H, T}} \right)^2 \left\{ \left(\overline{\Delta X}_{t_{\lfloor j/H \rfloor \cdot H, T}} \right)^2 - \left(\overline{\Delta X}_{t_j, T} \right)^2 \right\} \mid N_{\cdot,T} \right] \\ &= \frac{2T^{1/2}}{bH^2} \frac{1}{g_2^2} \sum_{i=1}^{N_T} \sum_{\alpha=1}^{H-1} K \left(\frac{t_i - u_o T}{bT} \right) K \left(\frac{t_{i+\alpha} - u_o T}{bT} \right) \mathbb{E} \left[\left(\overline{\Delta X}_{t_{\lfloor i/H \rfloor \cdot H, T}} \right)^2 \times \right. \\ &\quad \times \left. \left\{ \left(\overline{\Delta X}_{t_{\lfloor (i+\alpha)/H \rfloor \cdot H, T}} \right)^2 - \left(\overline{\Delta X}_{t_{i+\alpha}, T} \right)^2 \right\} \mid N_{\cdot,T} \right] \\ &= \frac{2T^{1/2}}{bH^2} \frac{1}{g_2^2} \sum_{i=1}^{N_T} K^2 \left(\frac{t_i - u_o T}{bT} \right) \mathbb{E} \left[\left(\overline{\Delta X}_{t_{\lfloor i/H \rfloor \cdot H, T}} \right)^2 \times \right. \\ &\quad \times \underbrace{\sum_{\alpha=1}^{H-1} \left\{ \left(\overline{\Delta X}_{t_{\lfloor (i+\alpha)/H \rfloor \cdot H, T}} \right)^2 - \left(\overline{\Delta X}_{t_{i+\alpha}, T} \right)^2 \right\}}_{(\clubsuit)} \mid N_{\cdot,T} \right] + o(1) \\ &= \frac{2T^{1/2}}{bH^2} \frac{1}{g_2^2} \sum_{i=1}^{N_T} K^2 \left(\frac{t_i - u_o T}{bT} \right) o \left(\frac{H^3}{T^2} \right) + o(1) = o_p(1), \end{aligned}$$

since g is differentiable, $g^{(1)}$ is Lipschitz's continuous, and

$$\begin{aligned}
(\clubsuit) &= \sum_{\alpha=1}^{H-1} \left(\overline{\Delta X}_{t_{\lfloor (i+\alpha)/H \rfloor \cdot H}, T} - \overline{\Delta X}_{t_{i+\alpha}, T} \right) \left(\overline{\Delta X}_{t_{\lfloor (i+\alpha)/H \rfloor \cdot H}, T} + \overline{\Delta X}_{t_{i+\alpha}, T} \right) \\
&= \sum_{\alpha=1}^{H-1} \sum_{l, l'=1}^{H-(j \bmod H)} \left\{ h \left(\frac{l + (j \bmod H)}{H} \right) - h \left(\frac{l}{H} \right) \right\} \left\{ h \left(\frac{l' + (j \bmod H)}{H} \right) - h \left(\frac{l'}{H} \right) \right\} \times \\
&\quad \times (X_{t_{\alpha+l}, T} - X_{t_{\alpha+l-1}, T}) (X_{t_{\alpha+l'}, T} - X_{t_{\alpha+l'-1}, T}).
\end{aligned}$$

Thus, we get $(\Delta_1) - (\Delta_2) = o_p(1)$. Analogously, $(\Delta_4) - (\Delta_3) = o_p(1)$ and hence $(T_1) = o_p(1)$. The rest terms (T_2) and (T_3) can be done in the same manner by employing $\overline{\Delta \varepsilon}_i = -\sum_{l=1}^{H-1} h(l/H) \varepsilon_{i+l}$ (also compare with (II) and (III)). Therefore we conclude that (A.7) = $o_p(1)$. \square

For the bias derivation of the tick-time volatility estimate $\hat{\sigma}_{avg}^2(\cdot)$ we will need the following results - Lemma 1 and Corollary 1. In fact, these results are investigated under a general setting for point processes allowing for stochastic intensity. More precisely, given a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_{t,T})_{t \in [0, T]}, \mathbb{P})$ a point process $N_{t,T}$ has an $\mathcal{F}_{t,T}$ -intensity $\lambda(t/T)$ if the conditions in Definition D7 Brémaud (1981) holds. For a stopping time τ we define $\mathcal{F}_{\tau, T}$ as consisting of sets $A \in \mathcal{F}$ for which $A \cap \{\tau \leq t\} \in \mathcal{F}_{t,T}$.

Lemma 1 Suppose the intensity process $\lambda(u)$ is bounded continuous and bounded away from zero uniformly in $u \in [0, 1]$, with probability 1. Then for $j \geq 0$ and $0 < l \leq 4$, it implies that

$$\begin{aligned}
\text{(i)} \quad &\mathbb{E} \left[\int_{t_i}^{t_{i+j}} \lambda(s/T) ds \mid \mathcal{F}_{t_i, T} \right] = j, \quad \text{(ii)} \quad \mathbb{E} \left[\left(\int_{t_i}^{t_{i+j}} \lambda(s/T) ds \right)^2 \mid \mathcal{F}_{t_i, T} \right] = j^2 + j \text{ and} \\
\text{(iii)} \quad &\mathbb{E} \left[(t_{i+j} - t_i)^l \mid \mathcal{F}_{t_i, T} \right] = O(j^l).
\end{aligned}$$

Proof. Since $M_{t,T} = N_{t,T} - \int_0^t \lambda(s/T) ds$ is a martingale and the arrival times t_i are stopping times, we get

$$j = \mathbb{E} [N_{t_{i+j}, T} - N_{t_i, T} \mid \mathcal{F}_{t_i, T}] = \mathbb{E} \left[\int_{t_i}^{t_{i+j}} \lambda(s/T) ds \mid \mathcal{F}_{t_i, T} \right]$$

by the optional sampling theorem. Moreover, it is clear that $\tilde{M}_{t,T} := M_{t,T}^2 - \int_0^t \lambda(s/T) ds$ is another martingale, thus

$$0 = \mathbb{E} [\tilde{M}_{t_{i+j}, T} - \tilde{M}_{t_i, T} \mid \mathcal{F}_{t_i, T}] = \mathbb{E} \left[\left(j - \int_{t_i}^{t_{i+j}} \lambda(l/T) dl \right)^2 \mid \mathcal{F}_{t_i, T} \right] - j,$$

i.e. $\mathbb{E} \left[\left(\int_{t_i}^{t_{i+j}} \lambda(s/T) ds \right)^2 \mid \mathcal{F}_{t_i, T} \right] = j^2 + j$. Therefore we have shown (i) and (ii). Burkholder-Davis-Gundy's inequality (Jacod and Protter 2012, p. 39) yields

$$\mathbb{E} [|M_{t_{i+j}, T} - M_{t_i, T}|^4 \mid \mathcal{F}_{t_i, T}] \leq C \cdot \mathbb{E} \left[\left(\int_{t_i}^{t_{i+j}} \lambda(s/T) ds \right)^2 \mid \mathcal{F}_{t_i, T} \right] = O(j^2).$$

By applying Hölder's inequality, $\mathbb{E} [|M_{t_{i+j}, T} - M_{t_i, T}|^3 \mid \mathcal{F}_{t_i, T}] = O(j^{3/2})$. Hence

$$\begin{aligned}
\mathbb{E} [(t_{i+j} - t_i)^4 \mid \mathcal{F}_{t_i, T}] &\leq C \cdot \mathbb{E} \left[\left(\int_{t_i}^{t_{i+j}} \lambda(s/T) ds \right)^4 \mid \mathcal{F}_{t_i, T} \right] \\
&= C \cdot \mathbb{E} [(M_{t_i, T} - M_{t_{i+j}, T} + j)^4 \mid \mathcal{F}_{t_i, T}] = O(j^4).
\end{aligned}$$

The rest can be then easily justified by employing Hölder's inequality. \square

Corollary 1 Let paths of $\lambda(\cdot)$ lie in $\mathcal{C}^{m', \gamma'}[0, 1]$ and satisfy Assumption 3.1 iii). For $j \in \mathbb{N}$,

$$\begin{aligned} \text{(i)} \quad \mathbb{E} \left[t_{i+j} - t_i \mid \mathcal{F}_{t_i, T} \right] &= \frac{j}{\lambda(t_i/T)} - \frac{1}{2} \frac{\lambda^{(1)}(t_i/T)}{\lambda^2(t_i/T)} \frac{j^2}{T} I_{\{m' \neq 0\}} + O \left(\frac{j^{1+m'+\gamma'}}{T^{m'+\gamma'}} I_{\{m' \neq 2\}} \right) + O \left(\frac{j^3}{T^2} I_{\{m'=2\}} \right), \\ \text{(ii)} \quad \mathbb{E} \left[(t_{i+j} - t_i)^2 \mid \mathcal{F}_{t_i, T} \right] &= \frac{j^2 + j}{\lambda^2(t_i/T)} + O \left(\frac{j^{2+\gamma'}}{T^{\gamma'}} I_{\{m'=0\}} + \frac{j^3}{T} I_{\{m' \neq 0\}} \right). \end{aligned}$$

Proof. By Lemma 1 we get

$$\mathbb{E} \left[\lambda(t_i/T)(t_{i+j} - t_i) \mid \mathcal{F}_{t_i, T} \right] = \underbrace{\mathbb{E} \left[\int_{t_i}^{t_{i+j}} \{ \lambda(t_i/T) - \lambda(l/T) \} dl \mid \mathcal{F}_{t_i, T} \right]}_{=:(\Lambda)} + j.$$

Therefore, the first statement (i) is verified by considering the following cases: (for $\omega \in \Omega$)

(a) for $\lambda(\cdot)(\omega) \in \mathcal{C}^{0, \gamma'}$,

$$(\Lambda) \leq C \cdot \left[\int_{t_i}^{t_{i+j}} \left(\frac{l - t_i}{T} \right)^{\gamma'} dl \mid \mathcal{F}_{t_i, T} \right] \leq C \cdot \frac{j^{1+\gamma'}}{T^{\gamma'}};$$

(b) for $\lambda(\cdot)(\omega) \in \mathcal{C}^{1, \gamma'}$,

$$\begin{aligned} (\Lambda) &= -\lambda^{(1)} \left(\frac{t_i}{T} \right) \mathbb{E} \left[\int_{t_i}^{t_{i+j}} \left(\frac{l - t_i}{T} \right) dl \mid \mathcal{F}_{t_i, T} \right] + O \left(\mathbb{E} \left[\int_{t_i}^{t_{i+j}} \left(\frac{l - t_i}{T} \right)^{1+\gamma'} dl \mid \mathcal{F}_{t_i, T} \right] \right) \\ &= -\frac{1}{2} \lambda^{(1)} \left(\frac{t_i}{T} \right) \mathbb{E} \left[\frac{(t_{i+j} - t_i)^2}{T} \mid \mathcal{F}_{t_i, T} \right] + O \left(\mathbb{E} \left[\frac{(t_{i+j} - t_i)^{2+\gamma'}}{T^{1+\gamma'}} \mid \mathcal{F}_{t_i, T} \right] \right) \\ &\stackrel{(*)}{=} -\frac{1}{2} \frac{\lambda^{(1)}(t_i/T)}{\lambda^2(t_i/T)} \frac{j^2}{T} + O \left(\frac{j^{2+\gamma'}}{T^{1+\gamma'}} \right); \end{aligned}$$

(c) for $\lambda(\cdot)(\omega) \in \mathcal{C}^{2, \gamma'}$,

$$\begin{aligned} (\Lambda) &= -\lambda^{(1)} \left(\frac{t_i}{T} \right) \mathbb{E} \left[\int_{t_i}^{t_{i+j}} \left(\frac{l - t_i}{T} \right) dl \mid \mathcal{F}_{t_i, T} \right] + O \left(\mathbb{E} \left[\int_{t_i}^{t_{i+j}} \left(\frac{l - t_i}{T} \right)^2 dl \mid \mathcal{F}_{t_i, T} \right] \right) \\ &= -\frac{1}{2} \lambda^{(1)} \left(\frac{t_i}{T} \right) \mathbb{E} \left[\frac{(t_{i+j} - t_i)^2}{T} \mid \mathcal{F}_{t_i, T} \right] + O \left(\mathbb{E} \left[\frac{(t_{i+j} - t_i)^3}{T^2} \mid \mathcal{F}_{t_i, T} \right] \right) \\ &\stackrel{(*)}{=} -\frac{1}{2} \frac{\lambda^{(1)}(t_i/T)}{\lambda^2(t_i/T)} \frac{j^2}{T} + O \left(\frac{j^3}{T^2} \right). \end{aligned}$$

To show (*) in (b) and (c), we have to apply the result (ii) beforehand. Without doing this we can approximate them to only the order $O(j^2/T)$; this order is, however, enough to show the assertion (ii). More precisely, similar to (a) - (c) without (*) we get

$$\begin{aligned} &\mathbb{E} \left[\left(\int_{t_i}^{t_{i+j}} \lambda(t_i/T) dl \right)^2 - \left(\int_{t_i}^{t_{i+j}} \lambda(l/T) dl \right)^2 \mid \mathcal{F}_{t_i, T} \right] \\ &= \mathbb{E} \left[\underbrace{\int_{t_i}^{t_{i+j}} \{ \lambda(t_i/T) - \lambda(l/T) \} dl}_{\text{see i) without (*)}} \cdot \underbrace{\int_{t_i}^{t_{i+j}} \{ \lambda(t_i/T) + \lambda(l/T) \} dl}_{=O(t_{i+j} - t_i)} \mid \mathcal{F}_{t_i, T} \right] \\ &= O \left(\frac{j^{2+\gamma'}}{T^{\gamma'}} I_{\{m'=0\}} + \frac{j^3}{T} I_{\{m' \neq 0\}} \right) \end{aligned}$$

by Lemma 1. In particular,

$$\mathbb{E} \left[\lambda^2 (t_i/T)(t_{i+j} - t_i)^2 \mid \mathcal{F}_{t_i, T} \right] = j^2 + j + O \left(\frac{j^{2+\gamma'}}{T^{\gamma'}} I_{\{m'=0\}} + \frac{j^3}{T} I_{\{m' \neq 0\}} \right).$$

□

Proof of Theorem 3.5. This proof is analogous to that of Theorem 3.4 with an exception of the bias term (IV), which is non-trivial in this case. Therefore, we have omitted its details and give only the outline of the proof. Let

$$\begin{aligned} \hat{\sigma}_{pre}^2(u_o) &:= \frac{T}{N} \frac{1}{g_2} \sum_{j=-M}^M k \left(\frac{j}{M} \right) \left(\overline{\Delta Y}_{t_{i_o+jH}, T} \right)^2 \\ &\quad - \frac{T}{2NH} \frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2} \sum_{i=-N}^N k \left(\frac{i}{N} \right) (Y_{t_{i_o+i}, T} - Y_{t_{i_o+i-1}, T})^2, \end{aligned}$$

where $M = M(T) = \lfloor N/H \rfloor$ (we might assume that $M = N/H$ is an integer). Similar to the previous theorem, we point out that the difference between the two statistics - $\hat{\sigma}_{pre}^2(u_o)$ and $\hat{\sigma}_{avg}^2(u_o)$ - is asymptotically negligible, i.e.

$$\sqrt{N/T^{1/2}} \{ \hat{\sigma}_{pre}^2(u_o) - \hat{\sigma}_{avg}^2(u_o) \} = o_p(1).$$

This means seeking the limit distribution of $\hat{\sigma}_{pre}^2(u_o)$ is sufficient to infer the limit of $\hat{\sigma}_{avg}^2(u_o)$. Likewise, we must show the following statements:

$$\begin{aligned} (I) : & \sqrt{\frac{N}{H}} \left(\frac{T}{N} \frac{1}{g_2} \sum_{j=-M}^M k \left(\frac{j}{M} \right) \left\{ \left(\overline{\Delta X}_{t_{i_o+jH}, T} \right)^2 - \mathbb{E} \left[\left(\overline{\Delta X}_{t_{i_o+jH}, T} \right)^2 \mid N, T \right] \right\} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \xi_A^2), \\ (II) : & \sqrt{\frac{NH}{T}} \left(\frac{T}{N} \frac{2}{g_2} \sum_{j=-M}^M k \left(\frac{j}{M} \right) \left(\overline{\Delta X}_{t_{i_o+jH}, T} \right) \left(\overline{\Delta \varepsilon}_{i_o+jH} \right) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \xi_B^2), \\ (III) : & \sqrt{\frac{NH^3}{T^2}} \left(\frac{T}{N} \frac{1}{g_2} \sum_{j=-M}^M k \left(\frac{j}{M} \right) \left\{ \left(\overline{\Delta \varepsilon}_{i_o+jH} \right)^2 - \mathbb{E} \left(\overline{\Delta \varepsilon}_{i_o+jH} \right)^2 \right\} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \xi_C^2), \end{aligned}$$

and the asymptotic biases

$$\begin{aligned} (IV) : & \sqrt{\frac{N}{H}} \left(\frac{T}{N} \frac{1}{g_2} \sum_{j=-M}^M k \left(\frac{j}{M} \right) \mathbb{E} \left[\left(\overline{\Delta X}_{t_{i_o+jH}, T} \right)^2 \mid N, T \right] - \sigma^2(u_o) \right) = o_p(1), \quad \text{and} \\ (V) : & \sqrt{\frac{NH^3}{T^2}} \left(\frac{T}{N} \frac{1}{g_2} \sum_{j=-M}^M k \left(\frac{j}{M} \right) \mathbb{E} \left[\left(\overline{\Delta \varepsilon}_{i_o+jH} \right)^2 \right] \right. \\ & \quad \left. - \frac{T}{2NH} \frac{\sum_{l=1}^{H-1} h^2(l/H)}{g_2} \sum_{i=-N}^N k \left(\frac{i}{N} \right) (Y_{t_{i_o+i}, T} - Y_{t_{i_o+i-1}, T})^2 \right) = o_p(1). \end{aligned}$$

As was pointed out, one of the most difficult parts in this proof is the derivation of (IV), which is not

obvious in the tick-time volatility estimation. More precisely, we have

$$\begin{aligned}
& \sqrt{\frac{N}{H}} \left\{ \frac{T}{N} \frac{1}{g_2} \sum_{j=-M}^M k\left(\frac{j}{M}\right) \mathbb{E} \left[\left(\overline{\Delta X}_{t_{i_o+jH}, T} \right)^2 \mid N, T \right] - \sigma^2(u_o) \right\} \\
&= \sqrt{\frac{N}{H}} \frac{T}{N} \frac{1}{g_2} \frac{1}{H} \sum_{j=-N}^N \left\{ k\left(\frac{\lfloor j/H \rfloor}{M}\right) - k\left(\frac{j}{N}\right) \right\} \mathbb{E} \left[\left(\overline{\Delta X}_{t_{i_o+\lfloor j/H \rfloor \cdot H}, T} \right)^2 \mid N, T \right] \\
&\quad + \sqrt{\frac{N}{H}} \left\{ \frac{T}{N} \frac{1}{g_2} \frac{1}{H} \sum_{j=-N}^N k\left(\frac{j}{N}\right) \mathbb{E} \left[\left(\overline{\Delta X}_{t_{i_o+\lfloor j/H \rfloor \cdot H}, T} \right)^2 \mid N, T \right] - \sigma^2(u_o) \right\} \\
&=: (i) + (ii),
\end{aligned}$$

where

$$(i) \leq C \cdot \sqrt{\frac{N}{H}} \frac{T}{N} \frac{1}{H} \sum_{j=-N}^N k^{(1)}(\dots) \left| \frac{\lfloor j/H \rfloor}{M} - \frac{j}{N} \right| \cdot \sum_{l=1}^{H-1} g^2\left(\frac{l}{H}\right) \frac{1}{T} = O\left(\sqrt{\frac{H}{N}}\right) = o(1)$$

by the boundedness of derivatives $k'(\cdot)$, and

$$\begin{aligned}
(ii) &= \sqrt{\frac{N}{H}} \left(\frac{1}{NHg_2} \sum_{j=-N}^N k\left(\frac{j}{N}\right) \sum_{l=1}^{H-1} g^2\left(\frac{l}{H}\right) \sigma^2\left(\frac{t_{i_o+\lfloor j/H \rfloor \cdot H+l}}{T}\right) - \sigma^2(u_o) \right) \\
&= O\left(\sqrt{\frac{N}{H}} \cdot \left|\frac{H}{T}\right|^\gamma\right) + \underbrace{\sqrt{\frac{N}{H}} \frac{1}{N} \sum_{j=-N}^N k\left(\frac{j}{N}\right) \left\{ \sigma^2\left(\frac{t_{i_o+j}}{T}\right) - \sigma^2(u_o) \right\}}_{(\clubsuit)} = o(1) + o_p(1),
\end{aligned}$$

as the segment condition holds. To obtain $(\clubsuit) = o_p(1)$, we need to apply Corollary 1 many times. Here we demonstrate (\clubsuit) only for the case $m, m' = 2$ to simplify notation (other cases are analogous). For $\sigma(\cdot) \in \mathcal{C}^{2,\gamma}$ and $\lambda(\cdot) \in \mathcal{C}^{2,\gamma'}$ we can expand

$$\begin{aligned}
& \sqrt{\frac{N}{H}} \left(\frac{1}{N} \sum_{j=-N}^N k\left(\frac{j}{N}\right) \left\{ \sigma^2\left(\frac{t_{i_o-j}}{T}\right) - \sigma^2\left(\frac{t_o}{T}\right) \right\} \right) \\
&= \sqrt{\frac{N}{H}} \frac{1}{N} \sum_{j=-N}^N k\left(\frac{j}{N}\right) \left((\sigma^2(u_o))^{(1)} \left(\frac{t_{i_o-j}}{T} - \frac{t_o}{T} \right) + \frac{(\sigma^2(u_o))^{(2)}}{2} \left(\frac{t_{i_o-j}}{T} - \frac{t_o}{T} \right)^2 + O\left(\left| \frac{t_{i_o-j}}{T} - \frac{t_o}{T} \right|^{2+\gamma}\right) \right) \\
&=: (A) + (B) + (C).
\end{aligned}$$

It is easy to see that

$$(A) = \sqrt{\frac{N}{H}} \frac{1}{N} (\sigma^2(u_o))^{(1)} \sum_{j=-N}^N k\left(\frac{j}{N}\right) \left(\frac{t_{i_o-j}}{T} - \frac{t_o}{T} \right) + o_p(1) =: (A_1) + o_p(1)$$

We will show that $\mathbb{E}[(A_1)^2] = o(1)$, which results that (A_1) is $o_p(1)$. By the symmetry of k we can rewrite

$$(A_1) = \sqrt{\frac{N}{H}} \frac{1}{N} (\sigma^2(u_o))^{(1)} \sum_{j=1}^N k\left(\frac{j}{N}\right) \left\{ \left(\frac{t_{i_o-j}}{T} - \frac{t_o}{T} \right) + \left(\frac{t_{i_o+j}}{T} - \frac{t_o}{T} \right) \right\},$$

thus (setting $\sigma' := (\sigma^2(u_o))^{(1)}$, $\sigma'' := (\sigma^2(u_o))^{(2)}$, $\lambda' := (\lambda(u_o))^{(1)}$ and $\lambda := \lambda(u_o)$)

$$\begin{aligned}\mathbb{E}[(A_1)^2] &= \frac{N}{H} \frac{1}{N^2} \sigma'^2 \sum_{i,j=1}^N k\left(\frac{i}{N}\right) k\left(\frac{j}{N}\right) \times \\ &\quad \times \left(\mathbb{E} \left[\left(\frac{t_{i_o-i}}{T} - \frac{t_{i_o}}{T} \right) \left(\frac{t_{i_o-j}}{T} - \frac{t_{i_o}}{T} \right) \right] + \mathbb{E} \left[\left(\frac{t_{i_o-i}}{T} - \frac{t_{i_o}}{T} \right) \left(\frac{t_{i_o+j}}{T} - \frac{t_{i_o}}{T} \right) \right] \right. \\ &\quad \left. + \mathbb{E} \left[\left(\frac{t_{i_o+i}}{T} - \frac{t_{i_o}}{T} \right) \left(\frac{t_{i_o-j}}{T} - \frac{t_{i_o}}{T} \right) \right] + \mathbb{E} \left[\left(\frac{t_{i_o+i}}{T} - \frac{t_{i_o}}{T} \right) \left(\frac{t_{i_o+j}}{T} - \frac{t_{i_o}}{T} \right) \right] \right) \\ &=: (I_1) + (I_2) + (I_3) + (I_4).\end{aligned}$$

Since $N_{\cdot,T}$ has independent increments, the non-overlapping interarrival times are independent, particularly $\mathbb{E}[(t_i - t_j)(t_k - t_l)] = \mathbb{E}[t_i - t_j] \mathbb{E}[t_k - t_l]$ for $l < k \leq j < i$. Therefore, by Corollary 1

$$\begin{aligned}(I_2) &= \frac{N}{H} \frac{1}{N^2} \sigma'^2 \sum_{i,j=1}^N k\left(\frac{i}{N}\right) k\left(\frac{j}{N}\right) \mathbb{E} \left[\frac{t_{i_o-i}}{T} - \frac{t_{i_o}}{T} \right] \mathbb{E} \left[\frac{t_{i_o+j}}{T} - \frac{t_{i_o}}{T} \right] \\ &= \frac{N}{H} \frac{1}{N^2 T^2} \sigma'^2 \sum_{i,j=1}^N k\left(\frac{i}{N}\right) k\left(\frac{j}{N}\right) \left\{ \frac{-i}{\lambda} - \frac{1}{2} \frac{\lambda'}{\lambda^2} \frac{i^2}{T} + O\left(\frac{i^3}{T^2}\right) \right\} \left\{ \frac{j}{\lambda} - \frac{1}{2} \frac{\lambda'}{\lambda^2} \frac{j^2}{T} + O\left(\frac{j^3}{T^2}\right) \right\} \\ &=: (I_{2,1}) + \dots + (I_{2,9})\end{aligned}$$

with

$$\begin{aligned}(I_{2,1}) &= -\frac{N^3}{HT^2} \cdot \frac{\sigma'^2}{\lambda^2} \left(\frac{1}{N} \sum_{i=1}^N k\left(\frac{i}{N}\right) \frac{i}{N} \right)^2, \\ (I_{2,2}) &= \frac{N^4}{HT^3} \cdot \frac{\sigma'^2}{2} \frac{\lambda'}{\lambda^3} \left(\frac{1}{N} \sum_{i=1}^N k\left(\frac{i}{N}\right) \frac{i}{N} \right) \left(\frac{1}{N} \sum_{j=1}^N k\left(\frac{j}{N}\right) \frac{j^2}{N^2} \right), \quad (I_{2,3}) = O\left(\frac{N^5}{HT^4}\right), \\ (I_{2,4}) &= -\frac{N^4}{HT^3} \cdot \frac{\sigma'^2}{2} \frac{\lambda'}{\lambda^3} \left(\frac{1}{N} \sum_{j=1}^N k\left(\frac{j}{N}\right) \frac{j}{N} \right) \left(\frac{1}{N} \sum_{i=1}^N k\left(\frac{i}{N}\right) \frac{i^2}{N^2} \right), \quad ((I_{2,2}) \text{ cancels out } (I_{2,4})) \\ (I_{2,5}) &= O\left(\frac{N^5}{HT^4}\right), \quad (I_{2,6}) = O\left(\frac{N^6}{HT^5}\right), \\ (I_{2,7}) &= O\left(\frac{N^5}{HT^4}\right), \quad (I_{2,8}) = O\left(\frac{N^6}{HT^5}\right) \text{ and } (I_{2,9}) = O\left(\frac{N^7}{HT^6}\right).\end{aligned}$$

In particular, we obtain $(I_2) = (I_{2,1}) + o(1)$ if $N^5/(HT^4) \rightarrow 0$ (which is satisfied by our segment conditions). We will see later that $(I_{2,1})$ is eliminated so that $(I_2) = o(1)$. Let us continue to (I_4) .

$$\begin{aligned}(I_4) &= \frac{N}{HT^2} \cdot \sigma'^2 \frac{1}{N^2} \sum_{i=1}^N k^2\left(\frac{i}{N}\right) \mathbb{E}[(t_{i_o+i} - t_{i_o})^2] \\ &\quad + \frac{N}{HT^2} \cdot \sigma'^2 \frac{1}{N^2} \sum_{i \neq j, j < i} k\left(\frac{i}{N}\right) k\left(\frac{j}{N}\right) \mathbb{E}[(t_{i_o+i} - t_{i_o})(t_{i_o+j} - t_{i_o})] \\ &\quad + \frac{N}{HT^2} \cdot \sigma'^2 \frac{1}{N^2} \sum_{i \neq j, j > i} k\left(\frac{i}{N}\right) k\left(\frac{j}{N}\right) \mathbb{E}[(t_{i_o+i} - t_{i_o})(t_{i_o+j} - t_{i_o})] \\ &=: (I_{4,1}) + (I_{4,2}) + (I_{4,3}),\end{aligned}$$

where

$$(I_{4,1}) = \frac{N}{HT^2} \cdot \sigma'^2 \frac{1}{N^2} \sum_{i=1}^N k^2\left(\frac{i}{N}\right) \left\{ \frac{i^2 + i}{\lambda^2} + O\left(\frac{i^3}{T}\right) \right\} \rightarrow 0 \quad (\text{by Corollary 1})$$

and

$$\begin{aligned}(I_{4,2}) &= \frac{N}{HT^2} \cdot \sigma'^2 \frac{1}{N^2} \sum_{i \neq j, j < i} k\left(\frac{i}{N}\right) k\left(\frac{j}{N}\right) \mathbb{E}[(t_{i_o+i} - t_{i_o+j})(t_{i_o+j} - t_{i_o}) + (t_{i_o+j} - t_{i_o})^2] \\ &=: (I_{4,2,1}) + (I_{4,2,2}).\end{aligned}$$

Again by the independence of interarrival times and Corollary 1 we get

$$\begin{aligned}(I_{4,2,1}) &= \frac{N}{HT^2} \sigma'^2 \frac{1}{N^2} \sum_{i \neq j, j < i} k\left(\frac{i}{N}\right) k\left(\frac{j}{N}\right) \{\mathbb{E}[t_{i_o+i} - t_{i_o}] \mathbb{E}[t_{i_o+j} - t_{i_o}] - (\mathbb{E}[t_{i_o+j} - t_{i_o}])^2\} \\ &= \frac{N^3}{HT^2} \frac{\sigma'^2}{\lambda^2} \frac{1}{N^2} \sum_{i \neq j, j < i} k\left(\frac{i}{N}\right) k\left(\frac{j}{N}\right) \frac{i}{N} \frac{j}{N} - \frac{N^3}{T^2} \frac{\sigma'^2}{\lambda^2} \frac{1}{N^2} \sum_{i \neq j, j < i} k\left(\frac{i}{N}\right) k\left(\frac{j}{N}\right) \frac{j^2}{N} + o(1) \\ &=: (I_{4,2,1,1}) + (I_{4,2,1,2}) + o(1)\end{aligned}$$

as $N^4/(HT^3) \rightarrow 0$ (in the same way as in the derivation of (I_2)). The next term

$$\begin{aligned}(I_{4,2,2}) &= \frac{N}{HT^2} \cdot \sigma'^2 \frac{1}{N^2} \sum_{i \neq j, j < i} k\left(\frac{i}{N}\right) k\left(\frac{j}{N}\right) \left\{ \frac{j^2 + j}{\lambda^2} + O\left(\frac{j^3}{T}\right) \right\} \\ &= \frac{N}{HT^2} \cdot \frac{\sigma'^2}{\lambda^2} \sum_{i \neq j, j < i} k\left(\frac{i}{N}\right) k\left(\frac{j}{N}\right) \frac{j^2}{N^2} + o(1)\end{aligned}$$

as $N^4/(HT^3) \rightarrow 0$. Evidently, the first term of $(I_{4,2,2})$ wipes out $(I_{4,2,1,2})$. Furthermore, by changing the role of i and j the same result can be derived for $(I_{4,3})$, therefore $(I_4) = (I_{4,2,1,1}) + (I_{4,3,1,1}) + o(1)$. Fortunately we see that the sum of $(I_{4,2,1,1})$ and $(I_{4,3,1,1})$ is exactly the negative of $(I_{2,1})$, thus

$$(I_2) + (I_4) = o(1).$$

Analogously, we can show that $(I_1) + (I_3) = o(1)$, so we can conclude that $(A1)$ is $o_p(1)$. Finally, the negligibilities of (B) and (C) is verified by $\mathbb{E}|(B)| = o(1)$ and $\mathbb{E}|(C)| = o(1)$. \square

Proof of Theorem 3.6. In the case of $m, m' = 0$, it is necessary to restrict the the bandwidth size \mathfrak{b} and the segment length N according to Theorems 3.3 and 3.5 respectively, i.e.

$$N/T^{1/2} = o(T^{\frac{\gamma}{1+2\gamma}}) \quad \text{and} \quad \mathfrak{b}T = o(T^{\frac{2\gamma'}{1+2\gamma'}})$$

in order to obtain those limit distributions. We analyze it by looking at the following situations:

(a) If $\gamma' = \frac{\gamma}{2\gamma+2}$ then $N/T^{1/2} = O(\mathfrak{b}T)$. This gives

$$\begin{aligned}& \sqrt{\mathfrak{b}T} \left\{ \widehat{\sigma}_{pavg}^2(u_o) \widehat{\lambda}(u_o) - \sigma^2(u_o) \lambda(u_o) \right\} \\ &= \underbrace{\widehat{\sigma}_{pavg}^2(u_o)}_{\xrightarrow{\mathbb{P}} \sigma^2(u_o)} \cdot \sqrt{\mathfrak{b}T} \left\{ \widehat{\lambda}(u_o) - \lambda(u_o) \right\} + \lambda(u_o) \underbrace{\frac{\sqrt{\mathfrak{b}T}}{\sqrt{N/T^{1/2}}}}_{=\sqrt{c_1}} \cdot \sqrt{N/T^{1/2}} \left\{ \widehat{\sigma}_{pavg}^2(u_o) - \sigma^2(u_o) \right\} \\ &\xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \sigma^4(u_o) \lambda(u_o) \int_{\mathbb{R}} \mathfrak{K}^2(x) dx + c_1 \lambda^2(u_o) \left\{ \delta \xi_A^2 + \frac{1}{\delta} \xi_B^2 + \frac{1}{\delta^3} \xi_C^2 \right\} \right),\end{aligned}$$

since the limits of the first and second terms are independent.

(b) If $\gamma' > \frac{\gamma}{2\gamma+2}$ then $N/T^{1/2} = o(\mathfrak{b}T)$. This leads to

$$\begin{aligned}\sqrt{N/T^{1/2}} \left\{ \widehat{\sigma}_{pavg}^2(u_o) \widehat{\lambda}(u_o) - \sigma^2(u_o) \lambda(u_o) \right\} &= \widehat{\lambda}(u_o) \cdot \sqrt{N/T^{1/2}} \left\{ \widehat{\sigma}_{pavg}^2(u_o) - \sigma^2(u_o) \right\} \\ &\xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \lambda^2(u_o) \left\{ \delta \xi_A^2 + \frac{1}{\delta} \xi_B^2 + \frac{1}{\delta^3} \xi_C^2 \right\} \right).\end{aligned}$$

(c) If $\gamma' < \frac{\gamma}{2\gamma+2}$ then $\mathfrak{b}T = o(N/T^{1/2})$, which implies that

$$\sqrt{\mathfrak{b}T} \left\{ \hat{\sigma}_{pavg}^2(u_o) \hat{\lambda}(u_o) - \sigma^2(u_o) \lambda(u_o) \right\} \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \sigma^4(u_o) \lambda(u_o) \int_{\mathbb{R}} \mathfrak{K}^2(x) dx \right).$$

Other cases ($m = 0$ together with $m' = 1, 2$; or $m = 1, 2$) can be verified by the same arguments with the corresponding restrictions on \mathfrak{b} and N . \square

REFERENCES

- Aït-Sahalia, Y., and Jacod, J. (2014), *High-Frequency Financial Econometrics*, Princeton, New Jersey: Princeton University Press.
- Ané, T., and Geman, H. (2000), “Order Flow, Transaction Clock, and Normality of Asset Returns,” *The Journal of Finance*, 55, 2259–2284.
- Bandi, F. M., and Russell, J. R. (2008), “Microstructure Noise, Realized Variance, and Optimal Sampling,” *The Review of Economic Studies*, 75, 339–369.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008), “Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise,” *Econometrica*, 76, 1481–1536.
- Belomestny, D. (2011), “Statistical Inference for Time-Changed Lévy Processes via Composite Characteristic Function Estimation,” *The Annals of Statistics*, 39, 2205–2242.
- Bibinger, M., Jirak, M., and Reiß, M. (2015), “Volatility Estimation Under One-sided Errors With Applications to Limit Order Books,” *ArXiv Preprint* [online], arXiv:1408.3768v4. Available at <http://arxiv.org/pdf/1408.3768v4.pdf>.
- Brémaud, P. (1981), *Point Processes and Queues*, New York: Springer-Verlag.
- Clark, P. K. (1973), “A Subordinated Stochastic Process Model With Finite Variance for Speculative Prices,” *Econometrica*, 41, 135–155.
- Dahlhaus, R. (1997), “Fitting Time Series Models to Nonstationary Processes,” *The Annals of Statistics*, 25, 1–37.
- Dahlhaus, R., and Neddermeyer, J. C. (2013), “Online Spot Volatility-Estimation and Decomposition with Nonlinear Market Microstructure Noise Models,” *Journal of Financial Econometrics*, 12, 174–212.
- Delbaen, F., and Schachermayer, W. (1994), “A General Version of the Fundamental Theorem of Asset Pricing,” *Mathematische Annalen*, 300, 463–520.
- Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H.E. (2003), “A Theory of Power-Law Distributions in Financial Market Fluctuations,” *Nature*, 423, 267–270.
- Griffin, J. E., and Oomen, R. C. (2008), “Sampling Returns for Realized Variance Calculations: Tick Time or Transaction Time?,” *Econometric Reviews*, 27, 230–253.
- Hall, P., and Heyde, C. C. (1980), *Martingale Limit theory and its Application*, New York: Academic Press.
- Hansen, P. R., and Lunde, A. (2006), “Realized Variance and Market Microstructure Noise,” *Journal of Business & Economic Statistics*, 24, 127–161.
- Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009), “Microstructure Noise in the Continuous Case: The Pre-Averaging Approach,” *Stochastic Processes and their Applications*, 119, 2249–2276.
- Jacod, J., and Protter, P. E. (2012), *Discretization of processes*. Berlin Heidelberg: Springer-Verlag.
- Jones, C. M., Kaul, G., and Lipson, M. L. (1994), “Information, Trading, and Volatility,” *Journal of Financial Economics*, 36, 127–154.
- Koo, B., and Linton, O. (2012), “Estimation of Semiparametric Locally Stationary Diffusion Models,” *Journal of Econometrics*, 170, 210–233.

- Kuo, H.-H. (2006), *Introduction to Stochastic Integration*, New York, Springer.
- Munk, A., and Schmidt-Hieber, J. (2010), “Lower Bounds for Volatility Estimation in Microstructure Noise Models,” *Borrowing Strength: Theory Powering Applications -A Festschrift for Lawrence D. Brown*, 6, 43–55.
- Plerou, V., Gopikrishnan, P., Gabaix, X., A Nunes Amaral, L., and Stanley, H.E. (2001), “Price Fluctuations, Market Activity and Trading Volume,” *Quantitative Finance*, 1, 262–269.
- Podolskij, M., and Vetter, M. (2009), “Estimation of Volatility Functionals in the Simultaneous Presence of Microstructure Noise and Jumps,” *Bernoulli*, 15, 634–658.
- Roueff, F., von Sachs, R., and Sansonnet, L. (2016), “Locally stationary Hawkes processes,” *Stochastic Processes and their Applications*, 126, 1710–1743.
- Reiss, M. (2011), “Asymptotic Equivalence for Inference on the Volatility From Noisy Observations,” *The Annals of Statistics*, 39, 772–802.
- Zhang, L. (2006), “Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach,” *Bernoulli*, 12, 1019–1043.
- Zhang, L., Mykland, P. A., and Ait-Sahalia, Y. (2005), “A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data,” *Journal of the American Statistical Association*, 100, 1394–1411.